Mechanistic and biocatalytic investigation of glutathione-dependent, ligninolytic enzymes

Von der Fakultät für Lebenswissenschaften

der Technischen Universität Carolo-Wilhelmina zu Braunschweig

zur Erlangung des Grades

einer Doktorin der Naturwissenschaften

(Dr. rer. nat.)

genehmigte

D i s s e r t a t i o n

von Jana Husarčíková
aus Ružomberok / Slowakei

1. Referentin: Professorin Dr. Anett Schallmey
2. Referent: Professor Dr. Dieter Jahn
eingereicht am: 15.10.2020
mündliche Prüfung (Disputation) am:17.12.2020
Druckjahr 2021

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch die Mentorin der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Publikationen

Jana Husarcikova, Hauke Voß, Pablo Domínguez de María and Anett Schallmey. Microbial β-etherases and glutathione lyases for lignin valorization in biorefineries: current state and future perspectives, Appl. Microbiol. Biotechnol. **2018**, *102*, 5391−5401. https://doi.org/10.1007/s00253-018-9040-3.

Jana Husarcikova and Anett Schallmey. Whole-cell cascade for the preparation of enantiopure β-*O*-4 aryl ether compounds with glutathione recycling, J. Biotechnol. **2019**, *293*, 1-7. https://doi.org/10.1016/j.jbiotec.2019.01.002.

Tagungsbeiträge

Jana Husarcikova and Anett Schallmey. Lignin valorization with GSH-dependent enzymes. 2018, 9th International Congress on Biocatalysis, Hamburg, Germany.

Posterbeiträge

Jana Husarcikova, Pere Picart, Hauke Voß and Anett Schallmey. Biomass valorization using GSH-dependent enzymes. 2016, The EMBO conference: "The biochemistry and chemistry of biocatalysis: From understanding to design", Oulu, Finland.

Jana Husarcikova, Hauke Voß and Anett Schallmey. Biomass valorization using glutathione-dependent enzymes. Biotrans 2017, 13th International Symposium on Biocatalysis and Biotransformations, Budapest, Hungary.

Jana Husarcikova and Anett Schallmey. Novel insights into β-etherase catalysis. 2018, 8th International Cebitec Research Conference, Bielefeld, Germany

Jana Husarcikova and Anett Schallmey. Novel insights into β-etherase catalysis. Biocat 2018, 9th International Congress on Biocatalysis, Hamburg, Germany.

Jana Husarcikova and Anett Schallmey. Whole-cell cascade using β-etherases and glutathione lyase with intracellular glutathione supply and recycling. Biotrans 2019, 14th International Symposium on Biocatalysis and Biotransformations, Groningen, Netherlands.

Studentenarbeit

The students Angelika Neitzel (2016), Jan Terbrack (2018) and Till Peters (2019) performed several experiments presented in this thesis as part of their Bachelor thesis projects.

ABSTRACT

The current demand for more environmentally benign processes requires the investigation of alternative renewable resources in different industry sectors to become independent from fossil fuels. Lignin, as a major polymer present in nature, contains significant amounts of aromatic building blocks and thus, is considered a valuable renewable feedstock for the future supply of aromatics. For that reason, the depolymerization and valorization of this commodity in biorefineries via various chemical and enzymatic approaches is of high importance. Several enzymes are known to act on lignin, but β-etherases together with glutathione lyases catalyze the selective cleavage of β-*O*-4 aryl ether bonds present in lignin in a highly stereoselective manner. These enzymes, belonging to the glutathione S-transferase superfamily, catalyze a reductive, glutathione-dependent ether bond cleavage.

In this thesis, a whole-cell catalyst combining β-etherases and glutathione lyase was studied in order to overcome the enzymes' glutathione dependence, which is regarded as a major limitation for their application in the production of aromatics from lignin polymer and its oligomers. Therefore, the whole-cell catalyst was investigated for efficient cleavage of the β-*O*-4 aryl ether bond in a lignin model substrate with intracellular GSH regeneration. This revealed that up to 5 mM of this substrate could be converted without addition of external GSH or carbon source. The applicability of the whole-cell process was successfully demonstrated in the kinetic resolution of the racemic lignin model substrate to provide enantiopure compounds.

Understanding the catalytic mechanism of an enzyme is one of the crucial requirements for its successful application. In that matter, a thorough analysis of the active sites of β-etherases LigE and LigF from *Sphingobium* sp. SYK-6 is presented, expanding our knowledge on the impact of different active-site residues on catalytic efficiency. By introducing mutations into the active sites of LigE and LigF, residues important for GSH-activation and catalysis have been identified. Additionally, multiple LigF mutants with significantly improved activity were obtained, representing an excellent starting point for further protein engineering.

Moreover, the substrate scope of β-etherases was expanded to non-lignin-derived, bisphenol A-related compounds in this thesis, demonstrating the potential of these enzymes for the degradation of other, man-made structures.

Table of contents

1 Introduction

1.1 Current societal challenges

The 21st century is defined by many social, economic and environmental challenges that humankind has to solve. Among many others, the biggest concerns are caused by increasing global population, poverty alleviation, environmental degradation, species extinction, and climate change. The problem of climate change is no longer just a scientific interest, but it progressively affects economics, sociology, law, national and local politics, and health [1–3]. In order to prevent the irreversible damage on the planet Earth, as reported by IPCC (The Intergovernmental Panel on Climate Change), some governments (including Germany) have made commitments to reduce greenhouse gas emissions to limit its destructive effects. To meet these commitments, aggressive policy action and a rapid shift away from fossil fuel use is necessary. For example, the German government recently introduced a major framework called “Climate Action Law” and the policy programme “Climate Action Programme 2030” that focus on climate action measures such as expansion of renewable resources, support for e-mobility, and the phaseout of coal-fired power. In order to achieve this courageous goal, alternative sources of gas, fuel and other chemicals derived from fossil fuel sources have to be developed. One of the promising alternatives is lignocellulose. Lignocellulosic material can be fully renewed/regrown on a very low time-scale, has often been considered a waste product, and avoids the controversy of the “food versus fuel” argument as seen in the first generation of biofuels. This makes lignocellulose a great candidate to become a valuable feedstock for transformation into desired products.

1.2 Lignin as a part of lignocellulose

Lignocellulosic biomass, an abundant and renewable resource, is a sophisticated and complex structural material of plant cell walls. Lignocellulose as a supramolecule consists of three main classes of biopolymers – cellulose, hemicellulose and lignin [4–6] – the combination of which provide structural stability and rigidity to the plant cell [7,8].

Cellulose, the most abundant polymer in the nature, is a polysaccharide made of β-1,4-linked D-glucose units and it accounts for 30-50 % of the dry weight of lignocellulose (depending on the biomass source). Currently industry’s most desired biopolymer, it is used in pulp and paper production and as a part of the biotechnological process for conversion of cellulose into fermentable sugars. These are utilized to produce

bioethanol, methane and even hydrogen [9].

Hemicellulose possesses a complex carbohydrate structure comprised of various polysaccharides, principally xylans (pentose) and mannans (hexose), which can be solubilized and extracted in various conditions (temperature, pH, moisture content) [4]. The function of hemicellulose in lignocellulosic biomass is to connect lignin and cellulose fibres (**Fig. 1.1**).

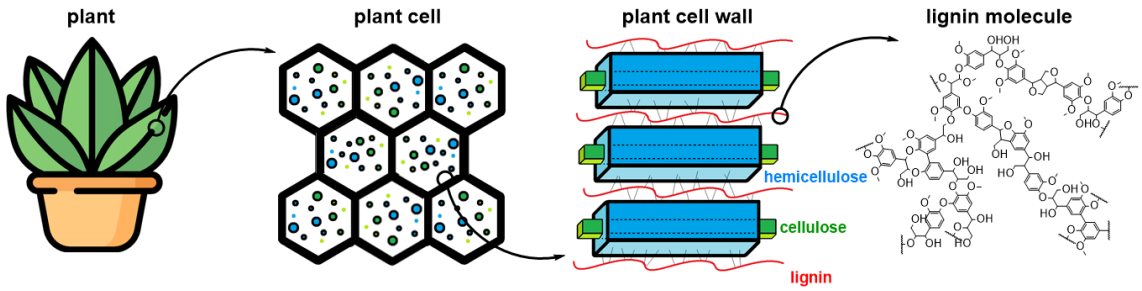


Fig. 1.1: Schematic representation of the location and structure of lignin in lignocellulosic material. Adapted from Zakzeski *et al.*[10].

Lignin, an irregular, three dimensional amorphous heteropolymer, is one of the most abundant polymers in nature [7,11]. 15-30 % of the dry weight of lignocellulose is comprised of lignin and its main function in the plant cell wall is to provide structural support, impermeability as well as resistance against oxidative stress and microbial attack [10]. All of these properties are directly connected with its structural features. Lignin consists of phenylpropanoid arylalcohol subunits held together by a minimum of 12 different described bonds [12]. The highly aromatic and functionalized matrix, together with the enormous quantities produced annually, make lignin an attractive raw material for the manufacturing of biofuels and fine as well as specialty chemicals. Although it has huge potential and offers attractive features, real industrial applications of lignin are still scarce due to its problematic processing caused by lignin's renowned recalcitrance and the variety of covalent inter-unit linkages through which its aromatic monomers are connected [13].

Lignin, currently mostly produced by pulp and paper industry as a by-product, is considered as a waste material and burned for the generation of heat and energy [10]. However, both lignin and lignocellulose are appealing feedstocks in biorefineries. In order to ensure environmentally-friendly, carbon efficient and renewable procedures in which biomass is converted into biofuel or chemicals, the understanding and controlling of the processing steps is crucial.

1.3 Lignin structure

Methoxylated hydroxycinnamyl alcohol building blocks are the major structural units of lignin. As opposed to cellulose with defined monomeric units linked by regular β-1,4-glycosidic bonds, lignin monomeric units (monolignols) are connected *via* a variety of non-enzymatic, radical coupling reactions that generate a mixture of interunit covalent bonds [14,15]. Generally, lignin is formed by radical polymerization of guaiacyl (G) units from precursor coniferyl alcohol, syringyl (S) units from precursor sinapyl alcohol, and *p*-hydroxyphenyl (H) units from precursor *p*-coumaryl alcohol [16,17] (**Fig. 1.2**). Depending on the biomass source, isolation method and environmental conditions, proportions between G:S:H units may vary. However, softwood lignin contains mostly G units and very low levels of H units, hardwood lignin contains similar levels of G and S units (and traces of H units), while grasses contain all three units [18]. In **Fig. 1.2**, the most common interunit linkages found in lignin are shown in a representative structure. Interunit linkages had been identified on low molecular weight products of lignin, monomers, dimers and oligomers, using GC-MS, LC-MS and NMR spectroscopy.

The most abundant linkage in lignin is the *arylglycerol-β-aryl ether* (β-*O*-4) bond. It corresponds to 40-60 % of all interunit linkages present in softwood as well as hardwood [14,19,20]. Recently, quantities of the β-*O*-4 units were measured directly in the cell wall using the two-dimensional NMR technique HSQC, which confirmed the predominant status of the β-*O*-4 linkage and its importance for lignin depolymerization [21].

The resinol linkage (β-β) is the core of resinol structures (pinoresinol, syringaresinol, episyringaresinol, lariciresinol, and dimethoxylariciresinol) [22], which were identified and isolated either chemically or enzymatically from hardwood lignin [22]. It was shown that hardwood lignin comprises larger amounts of resinol bonds than softwood and that maize lignin completely lacks resinol linkages in its structure [23].

Other lignin bonds, including the *phenylcoumaran bond* (β-5), *biphenyl bond* (5-5), *dibenzodioxocin bond* (α-β-*O*-4-4) and *diphenylether bond*, represent additional interlinkages that commonly prevail in the lignin structure [19,24,25]. A less common structure, the flavonoid tricin, has been recently confirmed to be present in the lignin structure of wheat straw [26,27] (**Fig. 1.2**). The valuable anticancer and antiaging properties of tricin may possibly increase the scientific desire to effectively cleave lignin linkages [28].

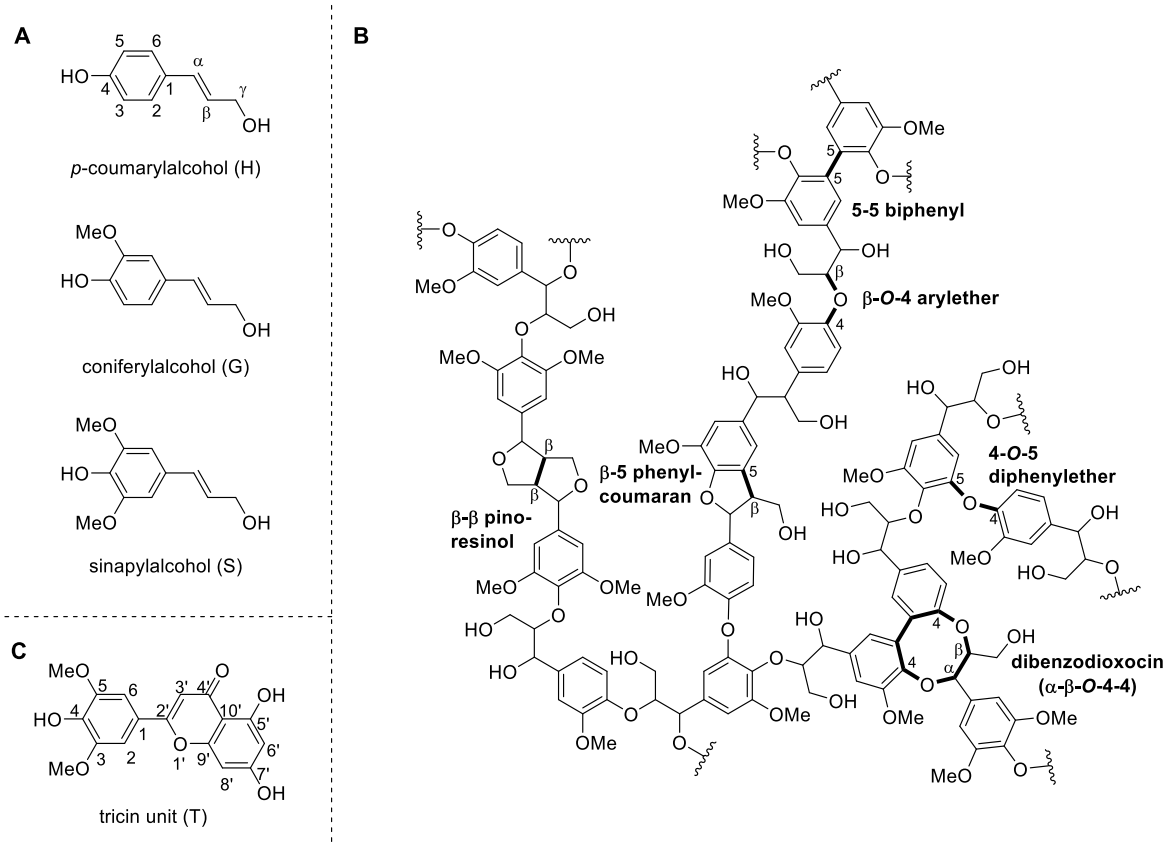


Fig. 1.2: Structure of lignin and its precursors. (**A**) Monolignols used in lignin synthesis in nature. (**B**) Representative structure of a lignin polymer with common linkage types highlighted in bold. (**C**) Flavonoid tricin.

1.4 Lignin isolation

As lignin covers 25-30 % of all renewable carbon in nature, numerous methods for the isolation of this polymer from its lignocellulosic matrix have been described. When lignin is isolated, different goals may be targeted and therefore different methods can be used. The general objectives for lignin isolations are: *analysis* (designed to determine lignin purity, native structure, or contend), *removal* (performed by the pulp and paper industry, designed to remove as much lignin from lignocellulose as possible), and *fractionation* (practised by biorefinery, designed for further conversion of lignin into biofuels/bio-based chemicals).

Klason lignin is one of the most broadly employed techniques used for quantifying lignin contend in biomass. Lignocellulosic feedstock is extracted by benzene and ethanol, then 72 % H_2SO_4 is added and the reaction is conducted at 20-30˚C for 1-4 h [29–31]. The result of this time and effort dependent process is acid-insoluble or ‘true’ lignin. Klason lignin is considered to be the most direct and the most reliable method for quantitative lignin analysis and is used as the standard method for lignin determination [32].

For the structural and compositional investigation (quality analysis), another ability of

lignin is employed. Solubility in organic solvents has been used for decades to extract lignin from biomass [33,34]. A variety of solvents can be used for lignin extraction such as 2-methoxyethanol, acetone, and dimethyl sulfoxide. However, due to its optimal Hildebrand solubility parameter (σ), dioxane is predominantly applied. The most commonly used method for lignin structure analysis is *Björkman lignin*. Described by Björkman in 1956 [35], the wood flour was exposed to extensive milling process in Lampén and vibrational ball mill in presence of toluene. Afterwards, milled wood lignin was purified from the aqueous *p*-dioxane. However, this lignin may contain some carbohydrate material, it is considered to be appropriate standard method resulting in lignin with low structural changes, and is used in structural and compositional studies [36]. Until now, most of the lignin production has been connected with pulp and paper manufacturing – historically the largest biomass refining industry. The pulp and paper industry's desired product is the cellulose pulp that is commonly obtained by either the kraft process or the lignosulfonate process [37,38]. During the *kraft lignin process*, wood biomass is treated with extensive amounts of aqueous sodium hydroxide and sodium sulfide at temperatures between 170-180˚C, allowing solubilization of most of the present lignin. While the result of such a process is a lignin-rich mixture called black liquor [39,40], most of it is not likely to serve as a source of lignin for biorefinery operations. Instead, the vast majority of concentrated black liquor is burned as a fuel to generate part of the industry's energy [17,41,42]. The reason for limited lignin valorization from black liquor are the structural changes in the lignin polymer that occur as a result of the kraft process. Compared to untreated lignin, kraft lignin usually contains more biphenyl units due to their resistance against heat and alkaline conditions during kraft process, contrary to other types of bonds that are degraded [43–46]. In addition, new types of functional groups and linkages are introduced during the pulping process, such as stilbenes [47] (**Fig. 1.3**).
Additionally to the kraft process, the pulp and paper industry uses a sulfite pulping, though on a smaller scale [39,48]. In the *sulfite process*, wood is usually treated with calcium or magnesium sulfite at 130-150˚C for several hours, allowing a sulfurous acid process to form benzylium cations. These cations attract sulfite ions and force the formation of lignosulfonic acid. In contrast to the kraft lignin, lignosulfonates are not largely burned as a source of energy, but are commonly used for further refinery [39]. The improved properties of the lignosufonates compared to kraft lignin (water solubility) empower further utilization of this feedstock, even though structural changes compared to the native lignin (incorporation of sulfonate groups on the arenes) are obvious (**Fig. 1.3**).

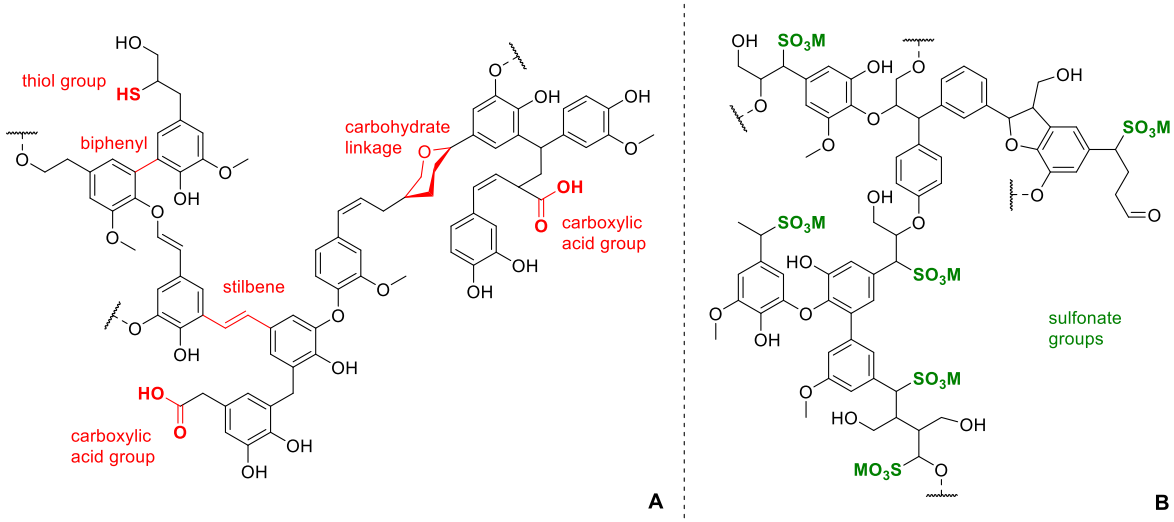


Fig. 1.3: Lignin model structure depicting features of kraft lignin (**A**) and ligninosulfonate lignin (**B**). Adapted from Zakzeski *et al*.[10].

As the availability of fossil fuels has its endpoint and is not renewable on any human time scale, different alternatives to fossil fuels, especially biogas, biodiesel and bioethanol, are in current demand. This, however, redefines the role of biomass and its derivatives in terms of availability, efficiency and the product spectra achieved from this raw material. Isolation of lignin, as it is performed in biorefinery, allows separation of the commodities of cellulose and hemicellulose as sources of fermentable sugars, while an isolated, structurally (relatively) unchanged lignin stream is open to further utilization into biobased chemicals.

Broadly used in biorefineries, the *Organosolv process* is a relatively general term describing the treatment of biomass with a broad range of organic solvents [40]. First described and patented by Kleinert & Tayenthal [267] for the fractionation of spruce, multiple variations of the process have been described by now [40,49–51]. Nonetheless, a significant majority of the organosolv procedures involve aqueous ethanol (or methanol) [52] as an essential organic solvent in combination with mineral acid as a catalyst [50,51]. Generally, the organosolv process consists of several steps, including hydrolysis of the hemicellulosic fraction, dissolution of lignin in the organosolv liquor and production of the cellulose in the form of pulp (after which the lignin can be recovered by precipitation or solvent evaporation) [53].

The organosolv process originated in the pulp and paper industry. However, it has transitioned into the more sophisticated biorefinery industry, the goal of which is to produce high quality lignin for further chemical production and simultaneous separation of the cellulose and hemicellulose. One of several currently applied processes in biorefineries is the *OrganoCat*® process, which provides separation of hemicellulose,

cellulosic pulp and lignin directly in one processing step [54]. The principle is based on the separation of the three main components in a biphasic system combining water and bio-based 2-methyltetrahydrofuran (2-MeTHF) as solvents as well as oxalic acid as a catalyst. Each part of the biomass is separated directly into the individual phases: the aqueous phase containing dissolved monomeric sugars from hemicellulose, the organic phase containing the extracted lignin fraction and the solid phase composed of the insoluble cellulose fraction [53,55] (**Fig. 1.4**). The ability to use a bio-based solvent, to recover the solvents and the catalyst, to separate the biomass in a single process step with a biomass loading of up to 100-150 g/L in the batch process and up to 400 g/L in repetitive batch, makes the OrganoCat® process not only ecologically but also economically feasible [53].

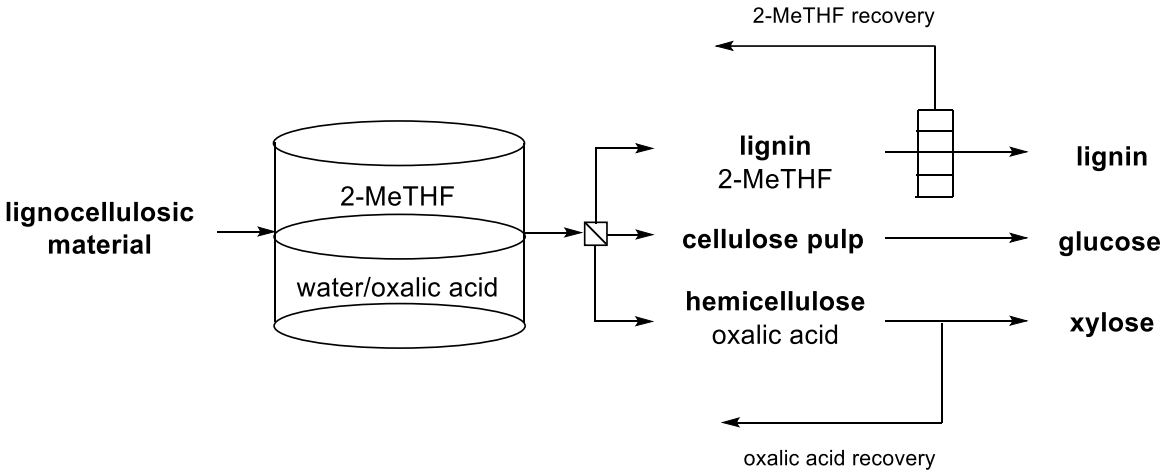


Fig. 1.4: The principle of the OrganoCat® process for the fractionation of lignocellulosic material to provide directly processible streams of hemicellulose, cellulose pulp, and lignin. Adapted from Grande *et al*. [53].

1.5 Lignin-degrading enzymes

As the industry of biorefineries is growing, methods for the valorization of produced high purity lignin is in current demand. The main thermo-chemical strategies for lignin depolymerization include pyrolysis, cracking or hydrolysis, hydrogenolysis, catalytic reduction, and catalytic oxidation [56]. These methods do not result in a selective lignin breakdown, but yield complex mixtures of degradation products. However, a selective cleavage of carbon-carbon and carbon-oxygen bonds in lignin is required for a more controlled and less complex product mixture. Enzymes that have evolved to depolymerize lignin may provide the solution for its successful, efficient and environmentally-friendly valorization. Enzymes capable of modifying the lignin structure have already been reported in the 1980's and over the years, various enzyme types and pathways for lignin degradation have been described [57,58].

1.5.1 Laccases (EC 1.10.3.2)

Laccases are polyphenol oxidoreductases belonging to the family of multi-copper oxidases (MCOs) found in plants, fungi, and bacteria. Although laccases are mostly extracellular enzymes, some wood-rotting fungi also produce intracellular laccases. They usually contain four copper ions, though atypical laccases with a reduced number of copper ions have been characterized [59]. The copper ions are divided into three types: a blue type-1 (T1) mononuclear copper center and a trinuclear copper cluster consisting of one type-2 (T2) or normal copper center and two type-3 (T3) or coupled binuclear copper centers. Type I copper is located in the T1 site of the enzyme and is responsible for the characteristic blue colour and start of the oxidation process in the laccase mechanism. The T2 copper, together with two ions of T3 copper, are located in a highly conserved trinuclear cluster [60]. Laccases are generally able to degrade phenolic substrates [61]. By combining them with natural mediators (acetovanillone, p-coumaric acid, ferulic acid, sinapic acid, syringaldehyde, acetosyringone, vanillin) or synthetic mediators (TEMPO (2,2,6,6-tetramethylpiperidine-1-oxyl), HBT (1-hydroxy-benzotriazole), ABTS (2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid), violuric acid), their substrate scope is extended to large, non-phenolic molecules, including lignin (**Fig. 1.5**) [62,63].

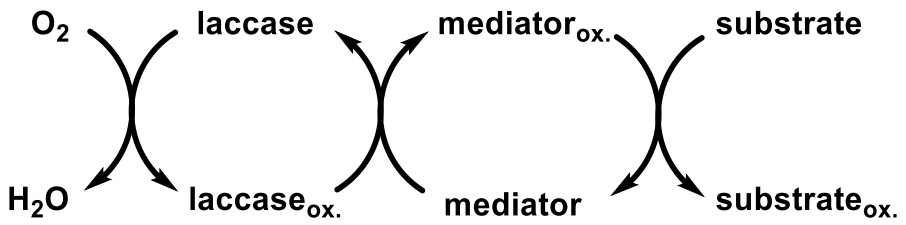


Fig. 1.5: Structural representation of laccase-mediator system, used in oxidation of variety of non-phenolic substrates.

Laccases catalyze a four single-electron substrate oxidation reaction and are fully reduced by O_2 through two consecutive two-electron transfer steps as shown in **Fig. 1.6** [64,65]. The first step of catalysis is the oxidation of the substrate by Cu^{2+} cations, forming a fully reduced enzyme. Reduction of dioxygen takes place in two steps via the formation of bound oxygen intermediates. The O_2 molecule binds to the T2-T3 site, resulting in the creation of a peroxide intermediate. The peroxide intermediate breaks down to an oxi-radical intermediate and undergoes a two-electron reductive cleavage of the oxygen-oxygen bond, releasing a water molecule and the resting enzyme in the substrate-binding state [59]. When laccases oxidize the substrate, free radicals are generated.

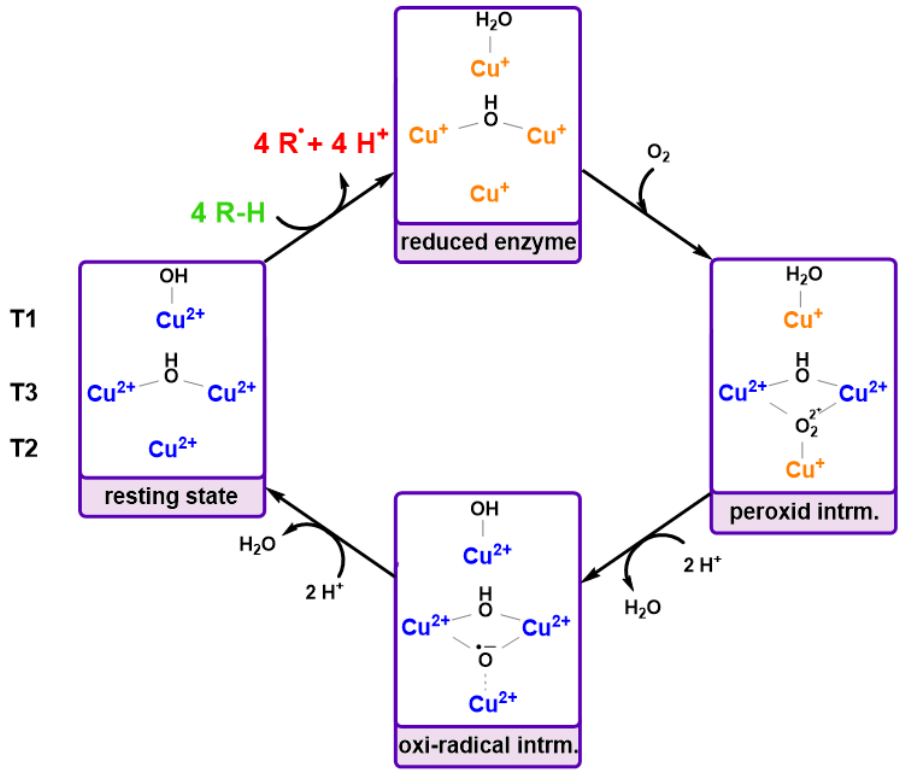


Fig. 1.6: Catalytic mechanism of multicopper laccases oxidizing substrate (green) to free radicals (red).

Laccases exist in a variety of structures; most of them are monomeric, but homodimeric, heterodimeric, and multimeric forms of laccases have been described [65,66]. The molecular weight of laccases depends on the source and ranges from 50 to 140 kDa. Structurally, laccases bear a cupredoxin-like fold arranged in a Greek-key β-barrel consisting of two β-sheets [67,68]. Due to their ability to detoxify a range of pollutants, laccases have been used in a variety of industrial and biotechnological applications (paper and pulp industry, textile industry, petrochemical industry, food industry, or cosmetic industry) [69].

1.5.2 Lignin peroxidases (LiP, EC 1.11.1.14)

LiPs, heme-containing glycoproteins of 38-46 kDa size, catalyze the H_2O_2-dependent oxidative depolymerization of lignin. These extracellular proteins are produced mostly by white-rot and brown-rot fungi, however some have been found in bacteria (actinomycetes, α-proteobacteria, and γ-proteobacteria) [4,70]. The high redox potential of LiPs caused by the heme iron allows oxidation of the aromatic non-phenolic structures of lignin [57,71]. Lignin polymer can only be degraded by LiPs with the help of natural or artificial mediators (as mentioned for laccases) as the polymer is too big to enter the

active site. The radical reaction mechanism of LiPs is similar to the mechanism of classical peroxidases (**Fig. 1.7**). One molecule of H_2O_2 oxidizes the resting state enzyme (with ferric heme) by removing two electrons and forming the most reactive intermediate, compound I. Compound I contains an oxo-porphyrin-radical (Fe^{4+}= O•) complex. Subsequently, it is reduced by one electron from a substrate/mediator (typically veratryl alcohol (VA)) generating compound II (Fe^{4+}= O, oxo-porphyrin complex). Compound II can be further reduced by another substrate/mediator molecule, returning the enzyme to its resting state. In the absence of substrate/mediator, a catalytically inactive form of the enzyme, compound lll, is formed. Compound lll is then converted to the resting enzyme by autooxidation or by oxidation with a VA radical [57,72].

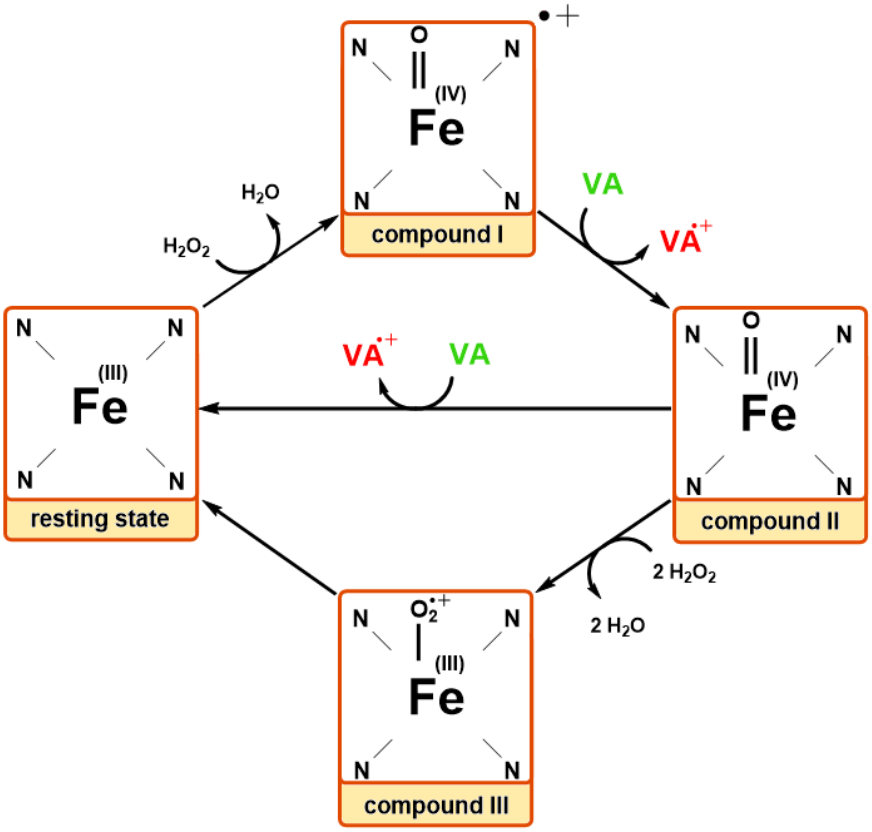


Fig. 1.7: Catalytic mechanism of lignin peroxidase oxidizing a substrate veratryl alcohol (VA, green) to free radicals (red).

Structurally, LiPs are composed of two domains (proximal and distal domain), in between which the heme is located. LiPs contain four disulfide bridges, two Ca^{2+} binding sites (one per each domain) and two glycosylation sites. Heterologous recombinant expression of the LiPs has been demonstrated in *E. coli* [73] and in the eukaryotic host *Pichia pastoris* [72]. Despite the fact that the LiPs are able to catalyze cleavage of a carbon-carbon bond in synthetic lignin, a drawback of the mechanism is represented by the re-

polymerization occurring due to the preferred radical coupling reaction with small phenolic products [74].

1.5.3 Manganese peroxidases (MnP, EC 1.11.1.13)

The most commonly produced peroxidases active on lignin are H_2O_2-dependent MnPs, which are found predominately in basidiomycetes (white-rot fungi, soil fungi), though they are produced by some bacteria and algae as well [57,75,76]. These extracellular, glycosylated heme-containing proteins, with a size of 40-50 kDa, are able to depolymerize and (re)polymerize a large number of phenolic substrates. MnPs are using the unique one-electron Mn^{2+}-dependent mechanism in which generated Mn^{3+} acts as diffusible redox-mediator, allowing for oxidation of variety of phenolic substrates like phenols, amines, dyes, as well as phenolic lignin model compounds. In contrast to the LiPs' mediator assisted mechanism, MnPs are not able to oxidize recalcitrant, non-phenolic compounds [77–79]. The catalytic cycle of MnPs shown in **Fig. 1.8** is initiated by binding of H_2O_2 to the native ferric enzyme and formation of compound I, which is again an oxo-porphyrin-radical (Fe^{4+}= O•) species. Subsequent reduction proceeds through a monochelated Mn^{2+}, that acts as one-electron donor and is oxidized to Mn^{3+}, creating compound II (Fe^{4+}= O, oxo-porphyrin intermediate). The reduction of compound II proceeds in a similar way and another Mn^{3+} is formed from Mn^{2+}, re-forming the native enzyme in the resting state [78,80–82].

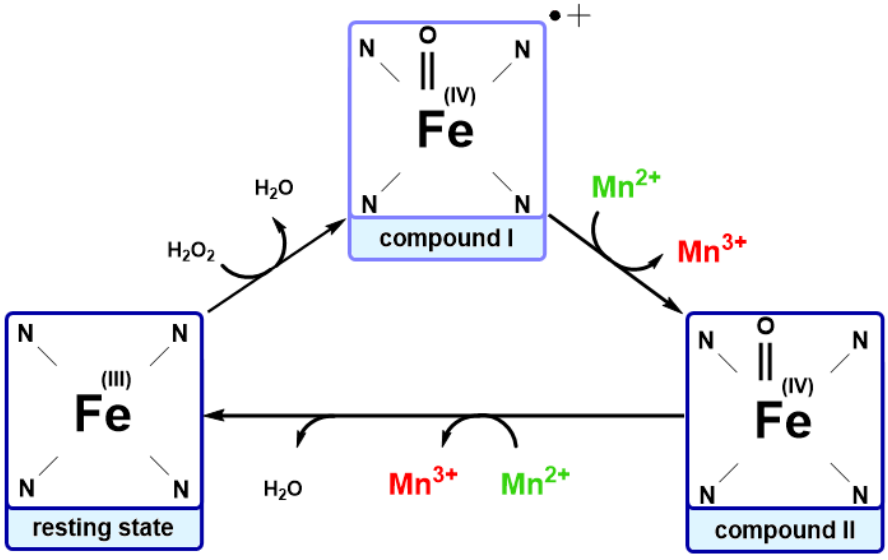


Fig. 1.8: Catalytic mechanism of manganese peroxidase via oxidation of Mn^{2+} (green) to Mn^{3+} (red).

LiPs and MnPs show similar structure: they are globular proteins formed by α-helices predominantly in two domains, with a central cavity containing the heme group. LiPs and MnPs stability depend on disulfide bonds: LiP contains four disulfide bridges, whereas MnP contains an additional bridge and two cation-binding sites that can accommodate a number of different metal ions.

1.5.4 Versatile peroxidases (VP, EC 1.11.1.16)

VPs are high redox potential, heme-containing enzymes widely studied for their unique characteristics. They are monomeric glycoproteins with a molecular weight of approximately 40-50 kDa and are exclusively expressed in white-rot fungi [83]. VP possess the ability to oxidize an extensive range of substrates while sharing structural and catalytic characteristics with LiPs and MnPs. Often referred to as 'hybrid' peroxidases, VPs are able to oxidize high-potential non-phenolic compounds like LiPs (Mn^{2+}-independent), but also are able to catalyze Mn^{2+}-mediated oxidations like MnPs. This catalytic versatility, together with no requirement for redox mediators, represents an attractive feature for oxidation of a variety of phenolic and non-phenolic aromatic compounds (including lignin). The catalytic cycle of VPs (**Fig. 1.9**) is initiated by binding of hydrogen peroxide to the ferric state of the heme, forming compound I, an oxo-porphyrin-radical (Fe^{4+}= O•) with significant oxidizing ability. Compound I is reduced to compound II, either via one-electron from a substrate/mediator or through a monochelated Mn^{2+}, that acts as one-electron donor and is oxidized to Mn^{3+}. The reduction of compound II continues in a comparable fashion either by another substrate/mediator molecule or by reduction of Mn^{2+} to Mn^{3+}, re-forming the native state of the enzyme.

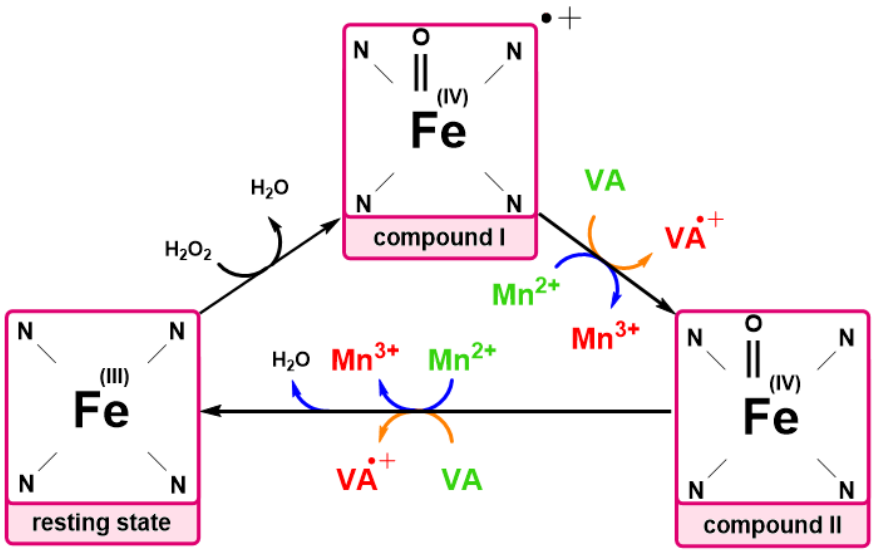


Fig. 1.9: Catalytic mechanism of versatile peroxidases via oxidation of a non-phenolic substrate veratryl alcohol (VA, green) forming free radicals (red) following the orange route (analog to LiPs mechanism), or via oxidation of Mn^{2+} (green) to Mn^{3+} (red) following the blue route (in analogy to the MnP mechanism).

VPs contain four conserved disulfide bridges, two structural Ca^{2+} sites, a heme pocket and a Mn^{2+}-binding site [84,85]. A high redox potential and the presence of different catalytic sites connected to the heme pocket are responsible for the promiscuity of these enzymes.

1.5.4 Dye-decolourizing peroxidases (DyP, EC 1.11.1.19)

The most recent group of heme peroxidases catalyzing H_2O_2-dependent oxidation of various molecules is named according to its unique ability to oxidize synthetic high-redox potential dyes of the anthraquinone type [86,87]. Among others, DyPs demonstrated oxidization of 2,2-azino-bis-(3-ethylbenzothiazoline-6-sulfonic acid) (ABTS), carotenoids, phenolic (DMP, guaiacol) and non-phenolic (VA) lignin models [88]. DyPs are mostly present in bacteria (actinobacteria and proteobacteria), but have been reported in fungi and higher eukaryotes as well [89]. Compared to the previously described lignin-modifying enzymes, DyPs show no structural homology to classic fungal heme peroxidases (LiP, MnP and VP). On the other hand, their reaction scope is similar to the typical peroxidases, LiPs, MnPs, and VPs, as they promote lignin degradation with the help of redox mediators. The catalytic mechanism of DyPs is suggested to be similar to that of LiPs including the resting state as well as transient compound I and compound II. The catalytic process typically requires two single-electron oxidations. In the first step, a resting ferric enzyme reacts with hydrogen peroxide forming the oxo-porphyrin-radical (Fe^{4+}= O•), compound I. This radical is then reduced by one electron to form a compound

II (Fe^{4+}= O, oxo-porphyrin-intermediate). Compound II undergoes a second reduction by one electron to restore the resting state of the peroxidase. In contrast to other heme peroxidases, our knowledge on the DyP mechanism is still limited [90]. Structurally, DyPs are rather different from the other peroxidases. Instead of distal histidine typical for most of the peroxidase, DyPs use an aspartame in their active site as acid-base catalyst. From the analyzed DyPs, it has been shown that they consist of two domains, each containing a ferredoxin-like fold built by predominantly antiparallel β-sheets [89,91,92], which is different from the α-helical fold of the other peroxidases. Heterologous expression of the DyPs from bacteria is commonly performed in *E. coli* cells, and the ability of DyPs to oxidize not only high redox substrates but also synthetic dyes represents a huge potential in biotechnological application of waste water treatment [93].

1.5.6 Glutathione S-transferases (GST; EC 2.5.1.18)

GSTs form a protein superfamily that plays an important role in cell detoxification. A shared character of all GSTs is their ability to catalyze conjugation of the tripeptide glutathione (GSH) to a wide range of hydrophobic substrates, and thus boost their excretion from cells. In nature, GSTs have been found in both eukaryotes and prokaryotes, and have been divided into four main families: cytosolic GSTs, mitochondrial GSTs, microsomal GSTs and bacterial fosfomycin-resistance proteins [94]. Each of the families have their classification system into classes named after greek letters. The amino acid sequences of the GST superfamily are extremely diverse and a large part of the sequences available in protein databases have not yet specified functions [94]. Structurally, GSTs often function as dimers; each monomer consists of a conserved N-terminal thioredoxin domain containing the GSH-binding site together with a more versatile C-terminal α-helical domain containing the binding site for the GSH acceptor.

1.5.6.1 GSH-dependent lignin degradation pathway

In 1989, Masai *et al*. [95] discovered an enzyme that was able to catalyze reductive cleavage of the β-*O*-4 aryl ether bond in lignin model compounds with a different mechanism when compared to the previously mentioned LiPs, MnPs, VPs, DyPs and laccases. Whereas these enzymes use an oxidative and non-selective radical-based mechanism when degrading lignin, the glutathione-dependent lignin degradation pathway is highly selective and follows a non-radical mechanism [96]. The first detailed studies on the GSH-dependent lignin degradation pathway were described for the proteobacterium *Sphingobium* sp. SYK-6 [95,97–100] (**Fig. 1.10**). This pathway consists of

three consecutive steps catalyzed by three sets of proteins. The first step of the cascade is represented by an NAD^+-dependent oxidation of the alcohol group at α carbon of the lignin substrate, forming the corresponding keto group. It involves four alcohol dehydrogenases, namely LigD, LigO, LigL, and LigN, which bear defined stereopreferences for the configuration at the α carbon atom of the lignin model substrate. While LigD and LigO display (*R*)-stereopreference, LigL and LigN possess affinity towards the opposite enantiomer, and exhibit (*S*)-stereopreference [101,102]. The produced ketone-intermediate is subsequently exposed to β-*O*-4 aryl ether cleavage catalyzed by highly enantioselective GSH-dependent β-etherases following an S_N2-type mechanism with inversion of stereoconfiguration at the β carbon [103,104]. From the three β-etherases that were described in *Sphingobium* sp.SYK-6, two, LigE and LigP, perform cleavage of ether bonds in substrates with (*R*)-configured β carbon, while the third, LigF, converts the corresponding (*S*)-enantiomers. The formed products, chiral glutathione adducts, are subsequently converted by GSH-dependent glutathione lyases catalyzing thioether cleavage and forming oxidized glutathione (GSSG) [104,105]. The complete outcome is a reductive cleavage of the β-*O*-4 aryl ether bond. Only one glutathione lyase, LigG, exhibiting high enantioselectivity for conversion of the (*R*)-configured thioether [103] was originally identified in *Sphingobium* sp. SYK-6, though low activity towards the corresponding (*S*)-enantiomer was described [105].

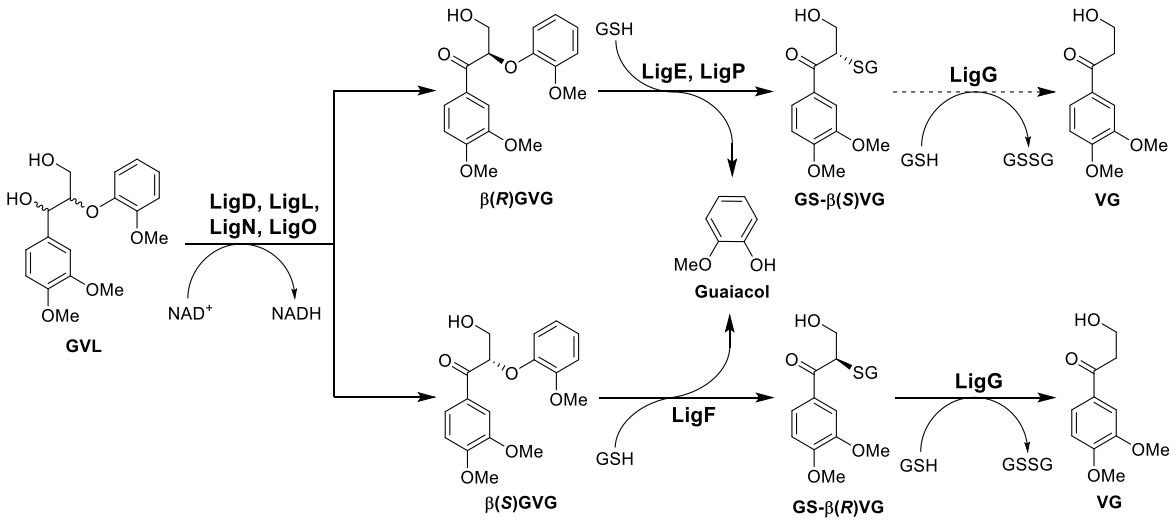


Fig. 1.10: Biochemical pathway of *Sphingobium* sp. SYK-6 for β-*O*-4 aryl ether cleavage [GVL: β-guaiacyl-α-veratrylglycerol, GVG: β- guaiacyl-α-veratrylglycerone, GS-VG: β-glutathionyl-α-veratrylglycerone, VG: β-deoxy-α-veratrylglycerone, GSH: reduced glutathione, GSSG: oxidized glutathione].

Alcohol dehydrogenases, β-etherases, and glutathione lyases in *Sphingobium* sp. SYK-6 are all intracellular enzymes, which means that they probably do not pursue polymeric lignin conversion, but rather degradation of smaller oligomeric products. On the contrary,

the only β-etherase that is secreted extracellularly and is GSH-independent has been described in an ascomycete of the genus *Chaetomium* [106]. In recent years, significant focus was given on the discovery of the new β-etherases and glutathione lyases in public databases. A variety of β-etherases was found and described in *Sphingobium* or *Novosphingobium* species [105,107–110] while recently, novel β-etherases from *Erythrobacter* and *Altererythrobacter* species were identified and characterized [111].

As previously described, heterogeneity, hydrophobicity and complex structure of lignin make its use and analysis in chemical or biochemical experiments rather challenging. Due to these issues, various lignin model compounds, oligomers with a defined structure, were developed. The benefit of these molecules is not only their representation of specific bonds found in lignin, but also more straightforward analysis with standard analytical methods. Since the β-*O*-4 aryl ether linkage is the most abundant bond in native lignin, many model compounds representing this link have been described and are shown in **Fig. 1.11**.

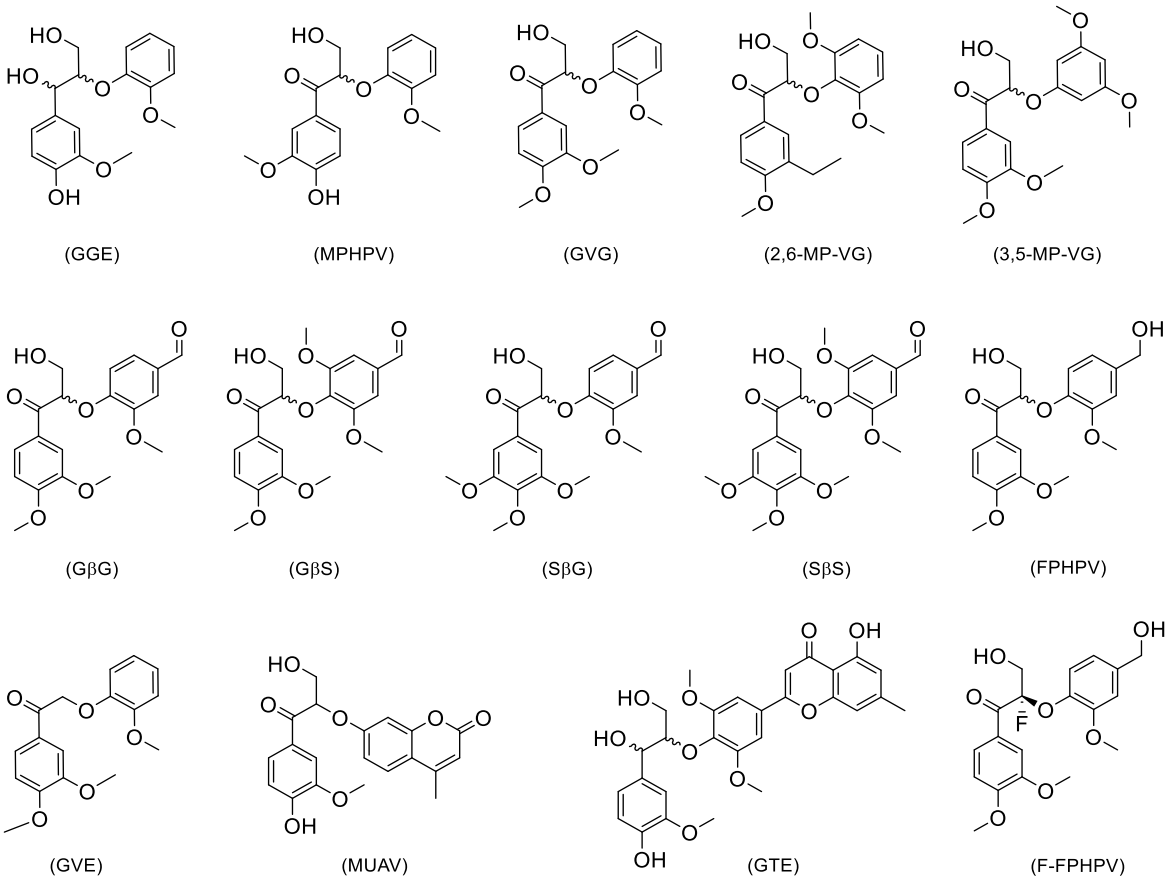


Fig. 1.11: Lignin-like model compounds commonly used to mimic the β-*O*-4 aryl ether bond found in lignin. [GGE: guaiacylglycerol-β-guaiacyl ether [110]; MPHPV: (2-methoxyphenoxy) hydroxypropiovanillone [108]; GVG: β-guaiacyl-α-veratrylglycerone [110]; 2,6-MP-VG: β-(2,6-methoxyphenoxy)-α-veratrylglycerone [110]; 3,5-MP-VG: β-(3,5-methoxyphenoxy)-α- veratrylglycerone [110]; GβG: guaiacyl-β-guaiacyl ether [107]; GβS: guaiacyl-β-syringyl ether [107]; SβG: syringyl-β-guaiacyl ether [107]; SβS: syringyl-β-syringyl ether [107]; FPHPV: β-(1-formyl-3- methoxyphenoxy)-γ-hydroxypropioveratrone [112]; GVE: β-guaiacyl-α-veratrylethanone [110]; MUAV: α-*O*-(β-methylumbelliferyl) acetovanillone [110]; GTE: guaiacylglycerol-β-tricin ether [27]; F-FPHPV: fluoro-(1-formyl-3-methoxyphenoxy)-γ-hydroxypropioveratrone [112].

1.5.6.2 β-Etherases

From all of the above-mentioned enzymes responsible for the GSH-dependent lignin degradation pathway, β-etherases have been studied the most. Besides the enzymes from *Sphingobium* sp. SYK-6, further bacterial, GSH-dependent etherases have been identified and biochemically characterized. Despite originating mostly from bacteria of the genus *Novosphingobium*, sequence similarities vary significantly. Sequence identities among LigE-type β-etherases vary between 56 % and 85 %, whereas LigF-type homologs share between 36 % and 96 % sequence identity, though the LigF homologs, based on the phylogenetic analyses, are split into two related but distinct subclades [109,111]. The special type of heterodimeric β-etherases was described in *N. aromaticivorans*, *Novosphingobium* sp. PP1Y and *Sphingobium xenophagum* by Kontur *et al*. [109] with sequence identities varying between 52 % and 74 %. On the other hand, novel sources of putative β-etherases with similar or even improved characteristics have

recently been described in *Erythrobacter*, *Altererythrobacter* and gammaproteobacteria with sequence identities among LigE- and LigF-type enzymes of 53 % and 54 %, respectively [111]. All studied β-etherases convert their substrates with absolute stereospecificity [107,108,110]. While all LigE-type homologs exhibit the same (*R*)-selectivity as LigE, all characterized members of the LigF subtype exhibit strict (*S*)-selectivity. The heterodimeric β-etherases, found in *N. aromaticivorans*, *Novo* sp. PP1Y and *Sphingobium xenophagum*, contrarily display stereoselectivity as the LigE-type homologs, even though their sequences suggest they are phylogenetically related to LigF-type enzymes [109]. Despite the fact that LigE- and LigF-type enzymes exhibit complementary stereoselectivities, their biochemical properties are rather similar. All described β-etherases possess pH optima in the alkaline pH range which supports deprotonation of the glutathione thiolate (pKa 9,65) for nucleophilic attack on the β carbon of the substrate. Yet, some β-etherases are able to retain their activity down to pH 5 [110]. The temperature optimum of β-etherases is usually around 20 to 30 °C, with some enzymes displaying a temperature optimum of even 40 °C [108]. Some were even able to retain activity at temperatures up to 60 °C [110], which together with their stereoselectivity gives them potential for practical applications in the kinetic resolution of desired chemicals [113] or lignin depolymerization within biorefineries. Crystal structures of β-etherases from *Sphingobium* sp. SYK-6, namely LigE and LigF were solved, which allows for a deeper understanding and characterization of the enzymes [112]. LigE and LigF, like other members of glutathione S-transferases, carry a N-terminal thioredoxin domain, consisting of four β strains and three α helices, connected via a small linker to the C-terminal helical domain. The position of the active site is located in between these domains, and while the active site of LigE forms a surface-exposed cleft, the active site of LigF is placed in a tunnel-like structure as shown in **Fig. 1.12**.

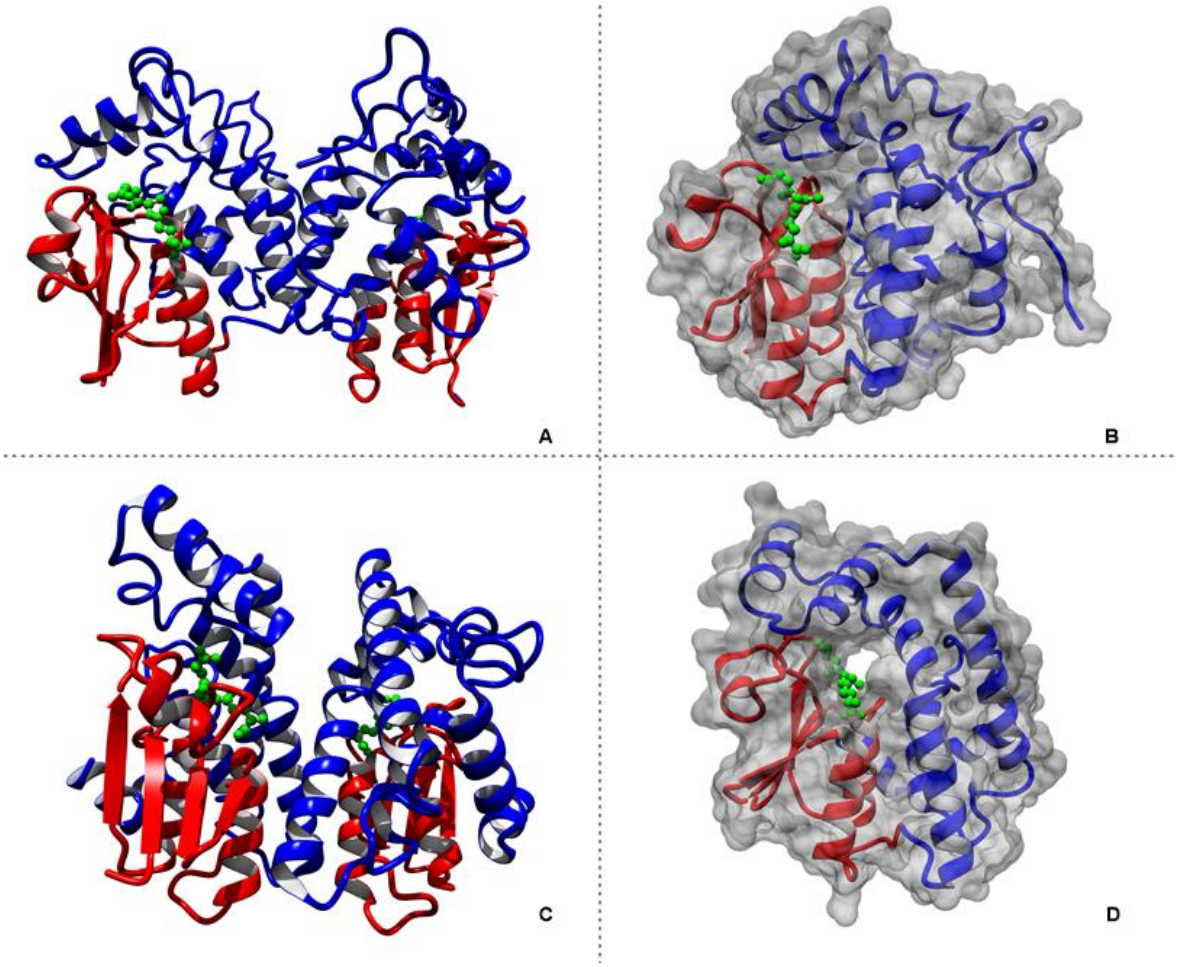


Fig. 1.12: Crystal structure of β-etherases LigE (A: dimer, B: monomer) and LigF (C: dimer, D: monomer). The co-crystallized cofactor GSH is visualized as green sticks and balls while the thioredoxin domain is shown in red and the helical domain is shown in blue.

The dimer interface of LigE is only formed by the C-terminal helical domain, whereas both thioredoxin and the helical domain contribute to the dimer formation in LigF. These structural differences are in agreement with the different classes of the GST superfamily they belong to. LigE belongs to the GSTFuA class, whereas classification of LigF is not yet clear due to its unique dimer interface [112,114]. Based on the crystal structures of β-etherases LigE and LigF with cofactor GSH bound to their active sites, it was possible to target potential amino acid residues important for catalysis. In both cases, a conserved serine residue (Ser21 in LigE and Ser13 in LigF) seems to play a role in catalysis since a major decrease in activity is observed when the residues are mutated [112]. Recently described heterodimeric β-etherases named BaeAB, suggest the importance of the Asn12 residue for the GSH activation necessary for the catalysis [115]. Based on the computational analysis of the LigF by Prates *et al*. [116], both hydrogen bonding residues, Ser13 and Asn12, have an activating impact on the GSH and thus on enzyme's catalysis.

1.5.6.3 Glutathione lyases

The last reaction step of the GSH-dependent lignin degradation pathway is performed by glutathione lyases. The further investigations showed that all β-etherase-containing strains also possess one or more glutathione lyases [105,109], suggesting that these enzymes usually work in close cooperation. Glutathione lyases catalyze the elimination of glutathione from the conjugate generated by the action of the β-etherases and thus allowing complete oxidation of the glutathione and the release of the final product. The only exception so far is represented by the glutathione lyase from the proteobacterium *Thiobacillus denitrificans.* Analysis of the genome of *T. denitrificans* revealed no presence of β-etherases genes, raising questions about the natural function of LigG-TD [105]. Interestingly, neither of the studied glutathione lyases displays absolute stereospecificity but rather a stereopreference, indicating the ability of glutathione lyases to convert both enantiomeric forms of the glutathione adduct, although conversion of one of the two enantiomers is always preferred [105,108]. So far, two glutathione lyases have been successfully crystalized, LigG from *Sphingobium* sp. SYK-6 [117,118] and $NaGST_{Nu}$ from *Novosphingobium aromaticivorans* [109]. Both, as members of the GST superfamily, possess a N-teminal thioredoxin domain and C-terminal helical domain. Regardless of their similar function, the enzymes are rather different structurally. LigG belongs to the Omega class of GSTs and exhibits a clear (*R*)-preference while $NaGST_{Nu}$ belongs to the Nu-class, which is less selective with slight (*S*)-preference [105,108,109]. In the structure of LigG, the active site forms a surface-exposed cleft, while the active site of $NaGST_{Nu}$ is located in a tunnel-like structure (**Fig. 1.13**). Interestingly, these structural properties together with the stereopreferences correspond to the respective characteristics of described β-etherases; LigG as well as LigE possess a surface-exposed active site and show (*R*)-selectivity/preference while both $NaGST_{Nu}$ and LigF contain the active site in a tunnel-like structure with an obvious (*S*)-selectivity/preference. This suggests that the shape of the active site plays a role in binding/accommodating the respective substrate enantiomer ((*R*)-enantiomer would not fit the to the tunnel-like active site).

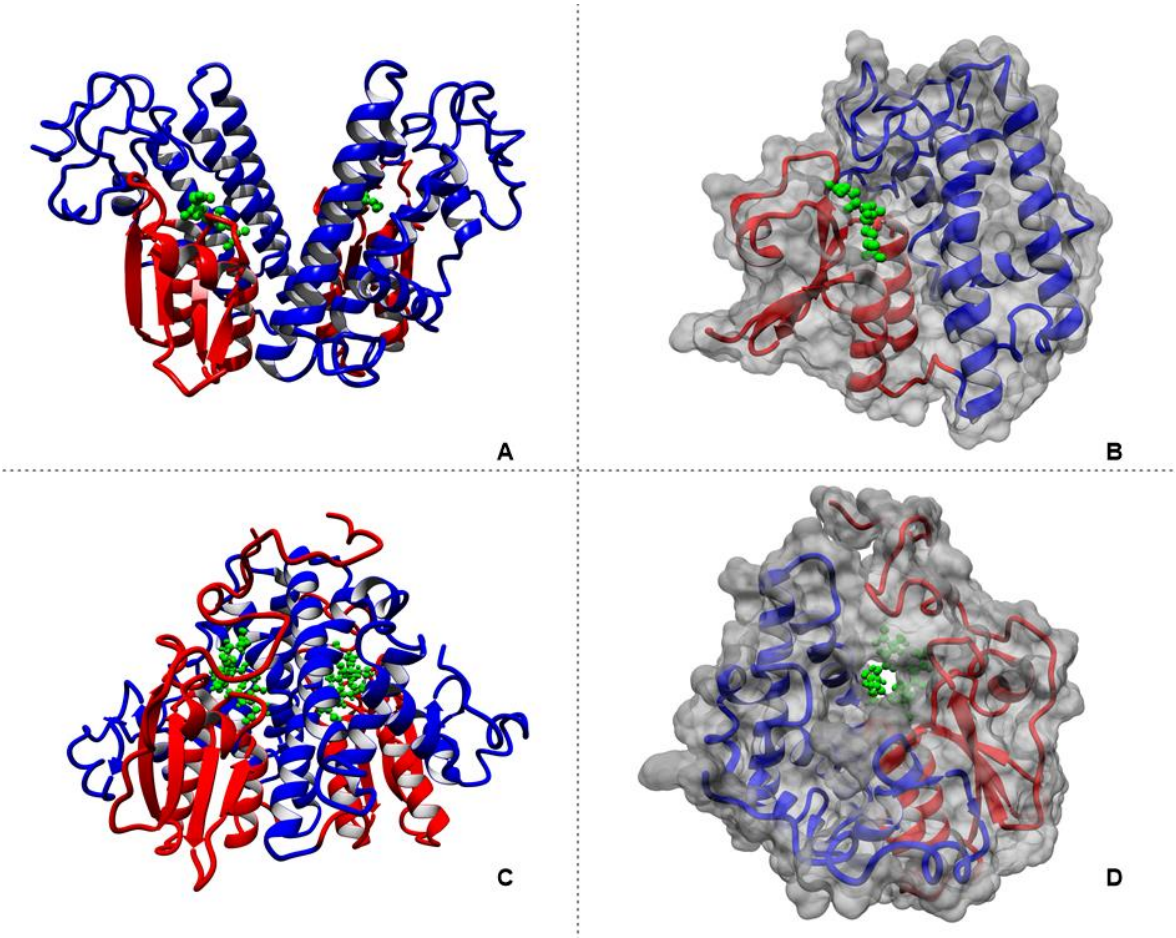


Fig. 1.13: Crystal structures of glutathione lyases LigG, co-crystallized with GSH (A: dimer, B: monomer), and NaGSTNu, co-crystalled with GSH (C: dimer, D: monomer). The co-crystallized cofactor GSH is visualized as green sticks and balls, while the thioredoxin domain is shown in red and the helical domain is shown in blue.

The crystal structure of LigG in complex with the substrate analog β-glutathionyl-acetoveratrone (GS-AV) was determined and analyzed by Pereira *et al*. [118]. Structural analysis of the positioning of GS-AV in the enzyme's active site allowed identification of amino acid residues important for catalysis. The position of the aromatic Tyr113 is important for interaction with the aromatic ring of the substrate as mutagenesis of this residue resulted in a drastic loss of activity. Based on the crystal structure, Pereira *et al*. [118] postulated a two-step mechanism for LigG catalysis. In the first step, the substrate analog, β-glutathionyl-acetoveratrone (GS-AV) is bound to the enzyme assisted by the interaction with Tyr113. The residue Cys15 simultaneously performs a nucleophilic attack on the carbon of the thioether bond of the substrate, resulting in formation of an enzyme-glutathione intermediate (via formation of a disulfide bridge) and release of the monolignan acetoveratrone (AV). In the second step, free GSH enters the active site and cleaves the enzyme-glutathione intermediate with formation and release of GSSG, restoring the native enzyme.

A different, single step mechanism was proposed by Kontur *et al*. [109] for Nu class glutathione lyases, such as $NaGST_{Nu}$ from *N. aromaticivorans*. Based on the crystal structure of $NaGST_{Nu}$, it was postulated that the substrate β-glutathionyl-hydroxypropiovanillone (GS-HPV) and GSH bind to the enzyme at the same time. The thiol group of GSH is activated by hydrogen bonds with Thr51 and Asn53 (part of the Thr-Pro-Asn motif conserved within Nu class GSTs) and nucleophilically attacks the thioether-carbon of the substrate to form GSSG, while hydroxypropiovanillone HPV is released.

1.5.6.4 Application of β-etherases and glutathione lyases in lignin depolymerization

β-Etherases and glutathione lyases are primarily characterized on lignin model structures, which allows a deeper understanding of the enzymes' characteristics. For possible industrial applications, however, natural, polymeric lignin is the desired starting material for the production of aromatic monomers. As discussed previously, the pre-treatment method by which the lignin is isolated has a serious impact on the distribution of carbon-carbon and carbon-oxygen bonds and hence, the efficiency of lignin depolymerization along with the type of released aromatics.

A first report for an *in vitro* biocascade combining LigD (Cα dehydrogenase), LigF (β-etherase) and LigG (glutathione lyase) from *Sphingobium* sp. SYK-6 for depolymerization of kraft lignin was described by Reiter *et al*. [119]. Combination of these enzymes in a one-pot reaction, together with the NADH-dependent glutathione reductase from *Allochromatium vinosum* (AvGR) for efficient NAD and GSH regeneration (**Fig. 1.14**), resulted in very low yields (~2 wt %) of different aromatics, mostly vanillin. The reason for this low efficiency may be a low residual content of the aryl ether linkages in reprocessed kraft lignin as well as the fact that the employed enzymes, due to their stereospecificity, were only able to cleave the (α*R* β*S*)-configured β-*O*-4 aryl ether bonds, corresponding to only 25 % of possible β-*O*-4 aryl ether linkages present in the lignin polymer.

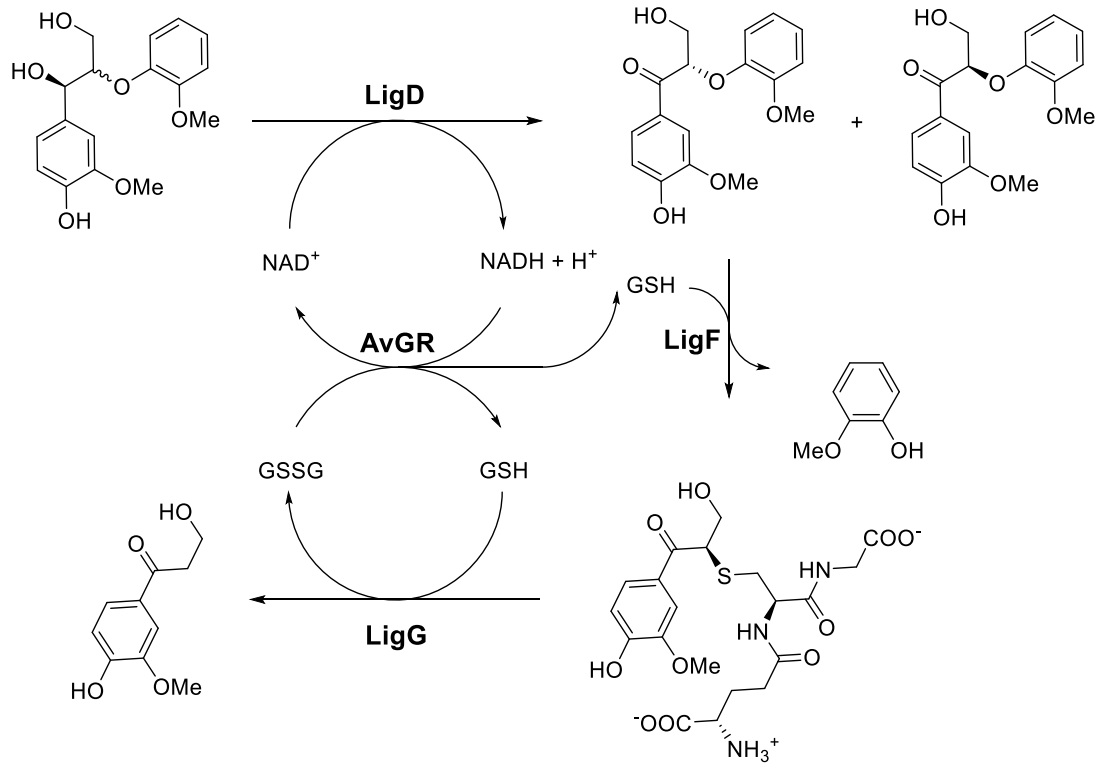


Fig. 1.14: Enzyme cascade for β-*O*-4 aryl ether bond cleavage. [LigD: Cα dehydrogenase, LigF: β-etherase, and LigG: glutathione lyase (all from *Sphingobium* sp. SYK-6), AvGR: glutathione reductase from *A. vinosum*]. Adapted from Reiter *et al*. [119].

This obstacle was overcome by Ohta *et al*. [120] when respective enzymes from *Novosphingobium* sp. MBES04 were combined, namely two stereocomplementary Cα-dehydrogenases (SDR3 and SDR5), two β-etherases with opposite stereoselectivity (GST4 and GST5) and a glutathione lyase (GST3) with a relatively low stereoselectivity. Depolymerization of milled wood lignin (MWL) from either softwood or hardwood sources using those enzymes in a one-pot system resulted in improved product yields (up to 6,6 wt %) of mostly guaiacyl-hydroxypropanone (GHP) and syringyl-hydroxypropanone (SHP) [120]. Improvement of the monomeric product yield may be explained by enantiocomplementary enzymes used in the process as well as by the structural properties of the milled wood lignin, which is reported to possess a lower degree of polymerization and facilitates easier substrate access to the enzyme's active site[121].

A very similar one-pot lignin depolymerization was performed by Gall *et al*. [27], applying LigD and LigN (enantiocomplementary Cα-dehydrogenases from *Sphingobium* sp. SYK-6), LigE and LigF (enantiocomplementary β-etherases from *Sphingobium* sp. SYK-6), the less selective glutathione lyase $NaGST_{Nu}$ from *N. aromaticivorans* and the glutathione reductase AvGR from *A. vinosum*. The mentioned cascade was tested to depolymerize a hybrid poplar (HP) lignin (enriched in syringyl units) as well as the structurally more diverse maize corn stover lignin. Here, different molecular weight (MW)

fractions of lignin were exposed to the enzymatic conversion. For lignin with an average MW of 1390 the highest monomeric product yield of up to 12.5 wt % was obtained, which was mostly composed of GHP and SHP [27]. The yield of retrieved monomeric products decreased with increasing size of the lignin oligomer, which can be explained by a lower solubility of the higher MW lignin fractions and the dimensional difficulties of the substrate to access the enzyme s' active sites.

Recently, a different approach was described by Picart *et al*. [122] in order to achieve oxidation of the α-hydroxyl groups in β-*O*-4 aryl ether linkages of lignin, which is necessary for subsequent cleavage of the ether bonds by β-etherase activity. Instead of stereocomplementary Cα dehydrogenases, a laccase-mediator system (LMS) was used. After unselective and cofactor-independent lignin oxidation by laccase lcc2 M3 from *T. versicolor* and violuric acid as mediator, cleavage of the β-*O*-4 aryl ether bonds was performed by β-etherases LigE from *Sphingobium* sp. SYK-6 and LigF-NA from *N. aromaticivorans*, along with glutathione lyase LigG-TD from *T. denitrificans* for glutathione removal. Using beech wood OrganoCat® lignin as a substrate, 12.5 wt % of a soluble lignin oil consisting of various low-molecular weight aromatics, particularly vanillin and coniferyaldehyde, was obtained. Due to the different required reaction conditions of laccase and GSH-dependent enzymes, this process was performed as a two-pot cascade with lignin oxidation as the first step followed by ether bond cleavage as the second step. A brief summary of all described lignin depolymerization strategies using β-etherases and glutathione lyases is shown in **Fig. 1.15**.

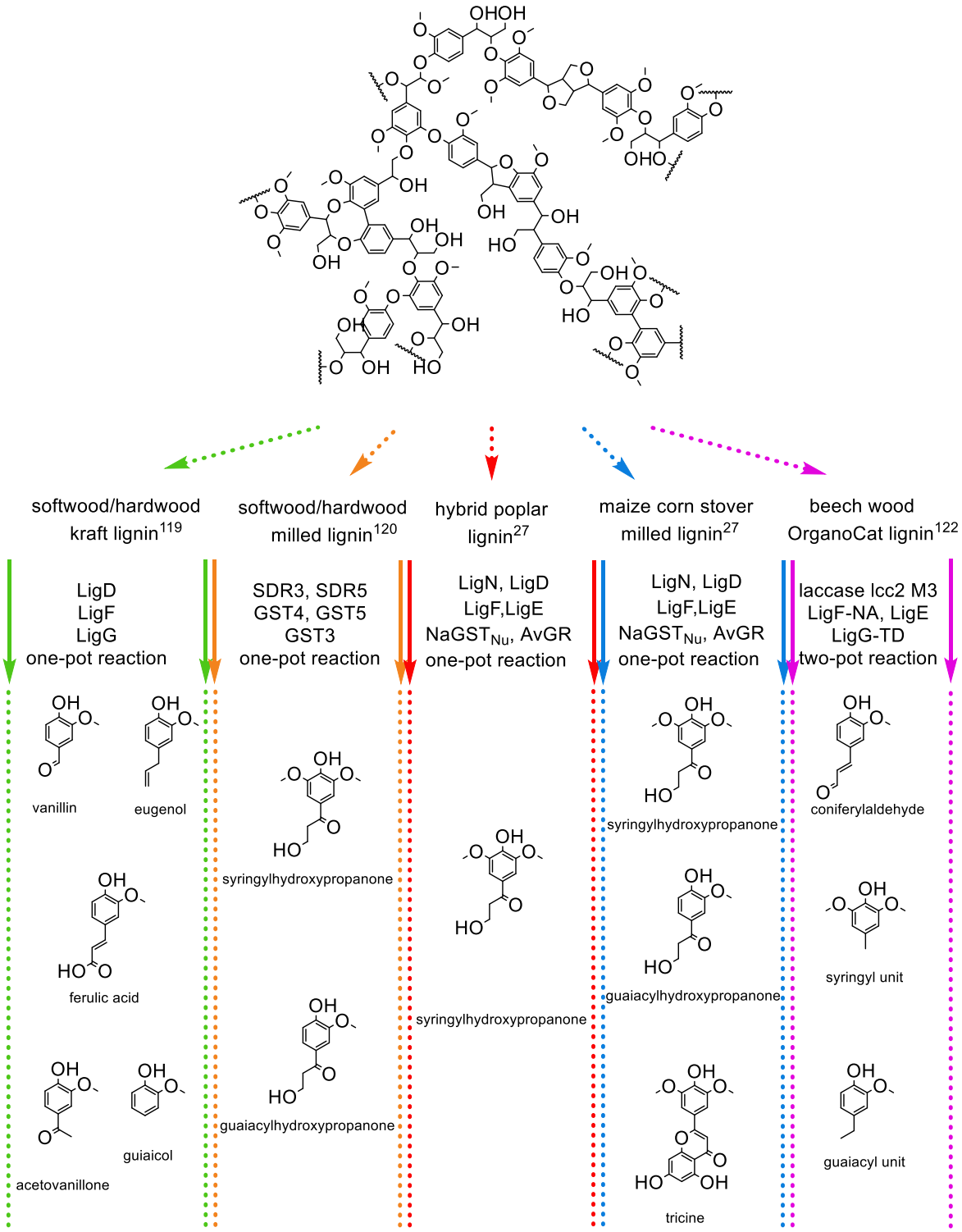


Fig. 1.15: Summary of enzymatic lignin depolymerization using β-etherases and glutathione lyases.

1.6 Protein engineering

Native enzymes are naturally optimized to catalyze chemical reactions within living organisms, which defines their optimal conditions. Mostly aqueous environment, neutral pH, ambient temperature and submillimolar concentrations of their natural substrates

may often represent obstacles in practical and industrial use. Therefore, most of the times, properties of native enzymes, such as activity, thermostability, selectivity, substrate concentration tolerance and solvent tolerance need to be adapted to fit the desired reaction conditions of a biocatalytic process. The structure and function of an enzyme are encoded in its amino acid sequence and therefore, modification of the sequence can lead to improved enzyme characteristics. Several methods have been described so far to effectively introduce mutations into the DNA sequence of an enzyme with the goal of developing improved enzyme variants suitable for industrial requirements. The DNA sequence of an enzyme can be exposed to several rounds of '*in vitro*' evolution until the desired characteristic is met. There are two main strategies of enzyme engineering: directed evolution and rational design.

1.6.1 Random mutagenesis and directed evolution

The definition of random mutagenesis lies in its name. Mutations are introduced into the DNA sequence (exchange of nucleotides) which can result in amino acid exchanges. When introducing random mutations into an enzyme-encoding gene, a mutant library is obtained that needs to be screened for the enzyme property of interest. Performing sequential cycles of random mutagenesis, in which the best enzyme variant from one cycle is taken as a template for the following cycle, is called directed evolution (**Fig. 1.16**). The pioneer of directed evolution, Francis Arnold (Nobel Laureate in 2018), was able to use only three rounds of random mutagenesis on the protease subtilisin E to remarkably improve its ability to function in a high concentration of DMF compared to the wild type enzyme [123]. Currently, several methods of directed evolution are commonly used: error-prone PCR, mutator strains or gene shuffling are just a few. The most popular technique is the error-prone PCR, which uses polymerases with a fairly high error rate (up to 2 %) together with unequal nucleotide concentrations and addition of manganese ions to increase the error rate of amplification of the wild-type sequence [124]. Among the biggest advantages of directed evolution is the ability to introduce mutations without deeper knowledge of the structure or mechanism of the enzyme of interest as well as the ability to run multiple sequential mutagenesis and screening cycles until the targeted enzyme improvement is met. On the other hand, a major obstacle of this method is the need for a fast, sensitive and reliable screening assay to identify improved mutants before the new round of mutagenesis is started [125].

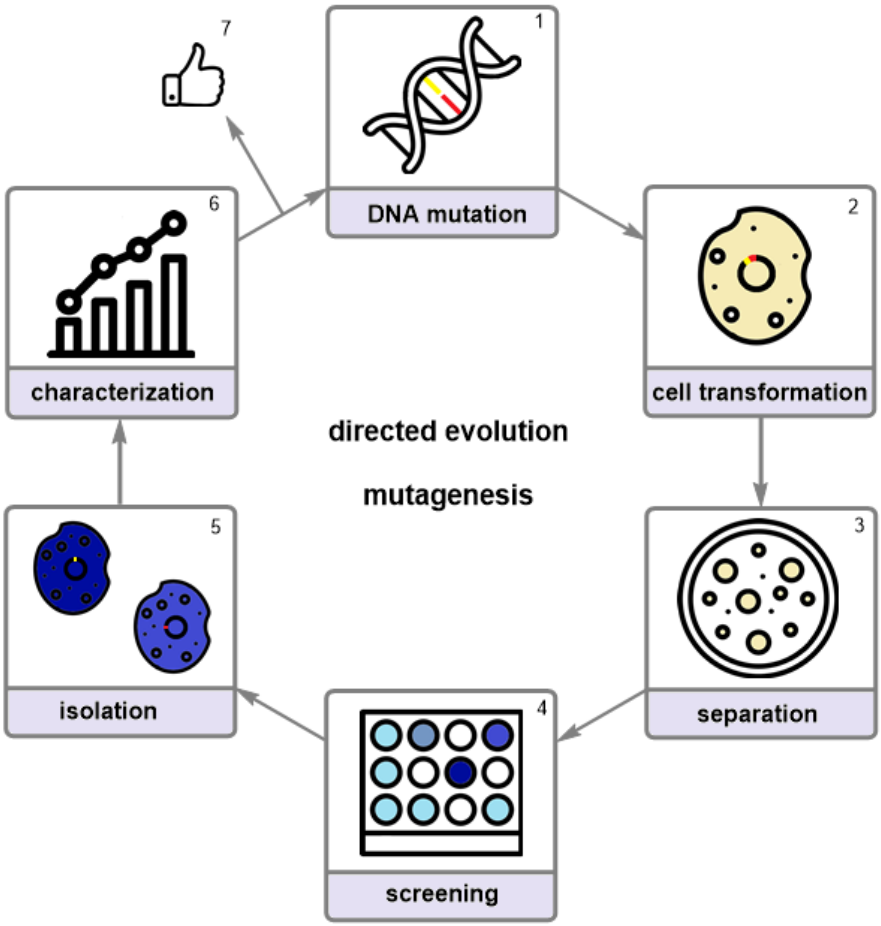


Fig. 1.16: General representation of directed evolution using high-throughput screening. 1: Random mutations are introduced to the whole gene or a gene fragment. 2: A bacterial cell is transformed with the mutated DNA. 3: Cell culture is grown on solid media, where individual cells (individual mutants) can be separated. 4: Screening of the generated mutant library to identify improved variants. 5: Improved variants are isolated and analyzed. 6: Selected mutants are characterized for desired properties. If an optimal mutant has been obtained, the enzyme evolution is finished and the process is completed (7). If not, the best variants of the previous mutagenesis round are used as a template for the next round, starting with random mutagenesis again (1), until the goal is met.

1.6.2 Rational design and semi-rational design

Contrary to directed evolution, rational design requires structural information about the enzyme to be engineered as well as detailed knowledge on its function and mechanism to ensure satisfactory results. In recent years, increased knowledge of protein structures due to protein NMR or X-ray crystallography made the rational design approach more and more popular among scientist. Together with computational analyses to predict the properties of each amino acid residue in the enzyme, it is possible to “spot” the residue or set of residues that may be important for an enzyme’s thermostability, substrate specificity or enantioselectivity [126]. This approach represents an efficiency improvement, as only few mutants are generated by rational design, and no screening method is required. Currently, plenty of commonly used methods are available on the market for

site directed mutagenesis, such as QuikChange or the Q5 protocol.

However, even with structural insights for an , it is difficult to predict the effect of an amino acid exchange with absolute certainty. Therefore, to overcome this problem a hybrid concept between rational design and random mutagenesis, called semi-rational design, has been developed. In this concept, one or more specific residues are targeted and randomly replaced by all (or multiple) amino acids. Although the choice of the position remains rational, the choice of the amino acids to be inserted is rather random. Site-saturation mutagenesis (SSM) represents one of the semi-rational approaches. In SSM, mutations are introduced into the DNA by using specific degenerated codons. Use of the degenerated “NNN” codon (N can be represented by all four nucleotides: A, T, G, and C) will allow for all 64 codons, and implements all 20 amino acids on this position. However, in this setting the screening of 192 random variants (4x4x4x3: four different nucleotides, three different positions in the codon, and an oversampling factor of three) is required to achieve 95 % theoretical library coverage [127]. The number of random mutants necessary to be screened increases dramatically if two residues are mutated at the same time (12 288 mutants). Therefore, smart libraries have been developed taking advantage of the repetition of the genetic code. By changing the degenerated codon to “NNS” (S coding for G or C) the number of codons is reduced to 32 while still covering all 20 amino acid residues. With this alteration, the number of random mutants to be screened decreases to 96 (4x4x2x3) for one mutated residue [125]. This concept of smart libraries was further advanced by using the “NDT” codon (D coding for A, G, or T), where the number of possible codons is reduced to 12 and the number of random mutants to be screened is reduced to only 36 (4x3x1x3) with a single mutated residue [128]. With this approach, however, “only” 12 different amino acids, containing representatives of aromatic, aliphatic, non-polar, polar, negatively and positively charged amino acid residues, are allowed on the selected position. On the other hand, it allows multiple residues to be mutated at the same time while keeping the number of mutants to be screened comparably low.

Semi-rational design accelerates the process of laboratory evolution of an enzyme and provides more in-depth characterization of the mutated position. In this matter, various smart library strategies to improve the enzyme’s efficiency, substrate acceptance, enantioselectivity, or thermostability have been described [127–129].

Every protein engineering approach has its benefits and its requirements. While random mutagenesis requires a quick and efficient screening method, the improvement of enzyme properties can be achieved even without structural knowledge of the enzyme. On the other hand, rational and semi-rational design depend on an extensive structural

and functional understanding of the enzyme, but put significantly less effort on the screening of mutants.

1.7 Aim of the project

The main focus of this thesis was a thorough investigation of β-etherases and glutathione lyases for efficient cleavage of β-*O*-4 aryl ether bonds to expand our knowledge on these enzymes for future practical applications.

As one of the drawbacks of these enzymes for practical application is their dependency on the relatively expensive cofactor glutathione, one aim of this thesis was to explore the possibility of intracellular provision and regeneration of glutathione by the *E. coli* metabolism. Since glutathione is produced as antioxidative agent by living cells, we wanted to investigate whether this can be utilized to supply the required glutathione cofactor for efficient intracellular β-*O*-4-aryl-ether-bond cleave in a lignin model substrate catalyzed by β-etherases and glutathione lyases. This concept was further applied to the kinetic resolution of a lignin model substrate on semi-preparative scale in order to afford enantiopure compounds.

Despite the fact that β-etherases are known for more than 30 years now, their catalytic mechanism has not been fully resolved yet. Based on the knowledge gained over the years, it is assumed that they follow an S_N2-like mechanism. In order to expand our knowledge on the catalytic mechanism of β-etherases, the second aim of this thesis was to study the impact of active site residues on LigE and LigF catalysis. This was approached by mutational studies of both enzymes, based on the previously published crystal structures, in combination with kinetic measurements.

Moreover, a third aim of this thesis was to investigate the potential ability of β-etherases to cleave also aryl ether bonds in non lignin-related substrates, which would significantly expand their biocatalytic applicability beyond lignin depolymerization.

2 Material and Methods

2.1 Materials

2.1.1 Chemicals

All chemicals were of analytical grade or higher and were purchased from either Sigma-Aldrich Chemie (Steinheim, Germany), VWR Chemicals (Darmstadt, Germany), Thermo Fisher Scientific (Waltham, USA), AppliChem (Darmstadt, Germany), Alfa Aesar (Karlsruhe, Germany), Acros Organics (Geel, Belgium) or Carl Roth (Karlsruhe, Germany) unless stated otherwise.

2.1.2 Equipment, software and kits

Tab. 2.1: Equipment used in this study.

Equipment	Supplier
Chromatography systems	
ÄKTA Pure protein purification system	GE Healthcare Bio-Science (Pittsburgh, USA)
ÄKTA Start protein purification system	GE Healthcare Bio-Science (Pittsburgh, USA)
HisTrap™ FF column	Agilent Technologies, Inc (Santa Clara, USA)
HisTrap™ HF column	Agilent Technologies, Inc (Santa Clara, USA)
Nexera XR HPLC System - LC-20AD XR - Sil-20 AC XR - SPD-M20A - CTO-20 AC - CBM-20A	Shimadzu Deutschland GmbH (Duisburg, DE)
Nucleosil® 100-5 C18 column (4.6 x 250 mm)	Macherey-Nagel (Düren, DE)
Chiralcel OD-RH column (4.6 × 150 mm)	Daicel Corporation (Illkirch, FR)
UV-Vis analytic systems	
CARY 60 Bio UV-Vis	Agilent Technologies, Inc (Santa Clara, USA)
CLARIOstar microtiter plate reader	BMG LABTECH GmbH (Ortenberg, DE)
NanoPhotometer NP80	Implen (Munich, DE)
Centrifuge	
Heraeus Fresco 21 microcentrifuge	Thermo Fisher Scientific (Waltham, USA)
Heraeus Multifuge X3R centrifuge	Thermo Fisher Scientific (Waltham, USA)
Mega Star 3.0 R centrifuge	VWR (Radnor, USA)
Micro Star 17 centrifuge	VWR (Radnor, USA)
Incubator	
INCU-Line Incubator	VWR (Radnor, USA)
MaxQ 8000 shaker	Thermo Fisher Scientific (Waltham, USA)
Minitron Infors HT	Infors AG (Bottmingen, CH)
ThermoMixer C	Eppendorf AH (Hamburg, DE)
VWR Incubating microplate shaker	VWR (Radnor, USA)
Electrophoresis	
FastGene blue/green GelPic LED Box	NIPPON Genetics Europe (Düren, DE)
Mini PROTEAN SDS-PAGE system	Biorad (Hercules, USA-CA)
Owl EasyCast B1A Mini Gel-electrophoresis systems	Thermo Fisher Scientific (Waltham, USA)
MS 300V power supply	Major Science (Saratoga, USA)

Scale	
Entris 224i-1S precision balances	Sartorius AG (Göttingen, DE)
Entris 3202i-1S precision balances	Sartorius AG (Göttingen, DE)
PCR devices	
PeqSTAR thermocycler	VWR (Radnor, USA)
Vacuum systems	
Rotavapor® R-300	BÜCHI Labortechnik Gmbh (Essen, DE)
RZ6 high vacuum pump	Vacuubrand (Wertheim, DE)
Other devices	
Arium® pro ultrapure water system	Sartorius AG (Göttingen, DE)
Eppendorf Research® plus pipettes	Eppendorf AH (Hamburg, DE)
Fisherbrand™Model 120 Sonic dismembrator	Thermo Fisher Scientific (Waltham, USA)
Sonorex Digitec	BANDELIN electronic Gmbh & Co. KG (Berlin, DE)
Vortex-Genie2	Scientific Industries Inc (Bohemia, USA)
Analytical devices used in cooperation	
Bruker Avance AVII 600 NMR spectrometer	Bruker Corporation (Billerica, USA)
LC MS - UHPLC system Ultimate3000RS - Kinetex C18 column (1,7µm 100A, 150x2,1mm) - Maxis HD UHR-TOF - Apollo II Elektrospray source	 Thermo Fisher Scientific (Waltham, USA) Phenomenex (Torrance, USA-CS) Bruker Corporation (Billerica, USA-MA) Bruker Corporation (Billerica, USA-MA)
MCS-ITC calorimeter	MicroCal (Northampton, USA)
Propol Digital Automatic Polarimeter	Dr. Kernchen (Seelze, DE)

Tab. 2.2: Software used in this study.

Software	Supplier
BioEdit 7.2	Tom Hall, Ibis Therapeutics (Carlsbad, USA)
BLASTP	NCBI (North Bethesda, USA)
ChemDraw Professional 16.0	Perkin Elmer (Waltham, USA)
Clone Manager 9.1	Scientific & Educational Software (Westminster, USA)
LabSolution 5.51	Shimadzu Deutschland GmbH (Duisburg, DE)
MestRenova LITE	Mestrelab Research, S.L. (Santiago de Compostela, ES)
MS Office	Microsoft Corporation ® (Redmond, USA)
OriginPro 2019b 9.65	OriginLab Corporation (Northampton, USA)
TopSpin 4.0.3	Bruker BioSpin GmbH (Ettlingen, DE)
UNICORN 5.31	GE Healthcare Bio-Science (Pittsburgh, USA)
Yasara	YASARA Biosciences GmbH (Vienna, AT)

Tab. 2.3: Kits used in this study.

Kit	Supplier
NucleoSpin® Gel and PCR Clean-up Kit	Macherey-Nagel (Düren, DE)
NucleoSpin® Plasmid Purification Kit	Macherey-Nagel (Düren, DE)
QIAquick® Gel Extraction Kit	Quiagen (Hilden, DE)
E.Z.N.A.® Plasmid Mini Kit I	Omega Bio-tek, Inc. (Norcross, USA)
E.Z.N.A.® MicroElute DNA CleanUp Kit	Omega Bio-tek, Inc. (Norcross, USA)
E.Z.N.A.® MicroElute Gel Extraction Kit	Omega Bio-tek, Inc. (Norcross, USA)
B-PER™ Bacterial Protein Extraction Reagent	Thermo Fisher Scientific (Waltham, USA)

2.1.3 Bacterial strains, plasmids and oligonucleotides

Tab. 2.4: Bacterial strains used in this study.

Strain	Characterization	Supplier
E. coli DH5α	F^- Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169 *rec*A1 *end*A1 *hsd*R17(r_k^-, m_k^+) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 λ^-	Thermo Fisher Scientific (Waltham, USA)
E. coli BL21 (DE3)	F^- *omp*T *hsdS*$_B$ (r_B^-, m_B^-) *gal dcm* (DE3)	Thermo Fisher Scientific (Waltham, USA)
E. coli BL21 Gold (DE3)	F^- *omp*T *hsdS*$_B$ (r_B^-, m_B^-) *gal dcm*$^+$ Tetr λ(DE3) *end*A *Hte*	Agilent Technologies (Santa Clara, USA)
E. coli C43 (DE3)	F^- *ompT hsdS*$_B$ (*r*$_B^-$, *m*$_B^-$) *gal dcm* (DE3)	Lucigen Corporation (Middleton, USA)

Escherichia coli DH5α was used for general cloning and plasmid DNA isolation, while *E. coli* BL21 (DE3), *E. coli* BL21 Gold (DE3) and *E. coli* C43 (DE3) were used for expression of the enzymes and the whole-cell biocatalytic reactions.

2.1.3.1 Strain constructs of the whole-cell experiments

The genes encoding LigF-NA, LigE and LigG-TD were previously cloned into vector pET-28a(+). To combine both etherase-encoding genes on a pETDuet-1 vector, the *ligE* gene was amplified by PCR using forward primer pET28a_NdeI_to_NcoI_fwd to introduce a NcoI restriction site and reverse primer pET28a_universal_rev (see **Tab. 2.6**). The resulting PCR product was digested with restriction enzymes NcoI and HindIII. The gene encoding LigF-NA was amplified by PCR using pET28a_universal_fwd and pET28a_universal_rev primers (**Tab. 2.6**) and the resulting PCR product was digested with NdeI and XhoI. Empty pETDuet-1 vector was first cut using restriction enzymes NcoI and HindIII, and the digested PCR product harboring the *ligE* gene was ligated into the first multiple cloning site (MCS). The resulting vector pETDuet-1*_ligE* was confirmed by sequencing. In a second step, vector pETDuet-1*_ligE* was digested with NdeI and XhoI and the digested PCR product harboring the LigF-NA gene was ligated into the second multiple cloning site. The resulting vector pETDuet-1*_ligE_ligF-NA* was again confirmed by sequencing. Furthermore, the gene encoding LigG-TD was amplified by PCR using pET28a_universal_fwd and pET28a_universal_rev primers (**Tab. 2.6**). The resulting PCR product was digested with NdeI and HindIII and ligated into empty pIT2 vector, which was cut with the same restriction enzymes. The resulting vector pIT2*_ligG-TD* was confirmed by sequencing. For co-expression of both β-etherases and the glutathione lyase, chemically competent *E. coli* BL21 (DE3) and *E. coli* C43 (DE3) strains were transformed with pETDuet-1*_ligE_ligF-NA* and pIT2*_ligG-TD*, resulting in strains *E. coli* BL21 (DE3) (pETDuet-1*_ligE_ ligF-NA*) (pIT2*_ligG-TD*) and *E. coli* C43 (DE3) (pETDuet-1*_ligE_ ligF-NA*) (pIT2*_ligG-TD*).

For co-expression of only one etherase and the glutathione lyase in *E. coli*, vector pIT2_*ligG-TD* together with pET-28a(+)_*ligE* or pET-28a(+)_*ligF-NA* was introduced into chemically competent *E.coli* C43 (DE3) cells, resulting in *E.coli* C43 (DE3) (pET28a(+)_*ligE*) (pIT2_*ligG-TD*) and *E.coli* C43 (DE3) (pET28a(+)_*ligF-NA*) (pIT2_*ligG-TD*) strains.

Tab. 2.5: Plasmids used in this study.

Plasmid	Characterization	Antibiotic resistance (working concentration *)	Source
pET-28a(+)	T7 promoter, T7 terminator, *lacI* coding sequence	Kanamycin (50 µg/mL)	Merck Group (Darmstadt, DE)
pETDuet-1	T7 promoter, T7 terminator, *lacI* coding sequence	Ampicillin (100 µg/mL)	Merck Group (Darmstadt, DE)
pIT2-MCS	*trc* promoter, rrnB terminator	Tetracycline (10 µg/mL**)	Kovach *et al.* [268]

*All of the antibiotics were prepared as stock solutions with 1000 times higher concentration.

**Tetracycline stock solution was prepared with 70 % ethanol.

Tab. 2.6: Oligonucleotides used in this study. The oligonucleotides were synthesized by Sigma-Aldrich Chemie (Steinheim, Germany).

Sequencing oligonucleotides	Sequence (5'→3')	Length (bp)	T_m (˚C)
pET28a_universal_fwd	GTGAGCGGATAACAATTCCC	20	63.2
pET28a_universal_rev	CTAGTTATTGCTCAGCGGTG	20	60.1
T7-fwd	TAATACGACTCACTATAGGG	20	53.2
T7-term	CTAGTTATTGCTCAGCGGT	19	54.5
pETUp	ATGCGTCCGGCGTAGA	16	54.3
DuetDown1	GATTATGCGGCCGTGTACAA	20	57.3
DuetUp2	TTGTACACGGCCGCATAATC	20	57.3
M13-rev	GAGCGGATAACAATTTCACACAGG	24	61.0
Altering oligonucleotides			
pET28a_NdeI_to_NcoI_fwd	GCAGCCCCATGGCACGCAATAATACCATC	29	79.5
Quikchange® oligonucleotides			
LigE-Y23A_fwd	CATATTTGGTACGCCAAACAGCCGGGCTAATGGTACAACCG	41	83.6
LigE-Y23A_rev	CGGTTGTACCATTAGCCCGGCTGTTTGGCGTACCAAATATG	41	83.6
LigE-W107A_fwd	CACGGACCAACTGCGGTTGAAGCCAGCCAATTATCCAGGAATT	43	86.5
LigE-W107A_rev	AATTCCTGGATAATTGGCTGGCTTCAACCGCAGTTGGTACGTG	43	86.5
LigE-F115A_fwd	CCGCAGTTGGTCCGTGGGCTCGTTGTTATATCCTGGA	37	85.1
LigE-F115A_rev	TCCAGGATATAACAACGAGCCCACGGACCAACTGCGG	37	85.1
LigE-Y122A_fwd	GCGGCAGGCTCAGATCATGAGCATCCAGGATATAACAACGA	41	84.8
LigE-Y122A_rev	TCGTTGTTATATCCTGGATGCTCATGATCTGAGCCTGCCGC	41	84.8
LigE-F142A_fwd	CGCTGACCTCCCAGAGCCCACTGTTCACGGC	31	85.8
LigE-F142A_rev	GCCGTGAACAGTGGGCTCTGGGAGGTCAGCG	31	85.8
LigE-W197A_fwd	GCAACGCTTGCGGTCGCCAGAAAAACTGCCAGTG	34	86.4
LigE-W197A_rev	CACTGGCAGTTTTTCTGGCGACCGCAAGCGTTGC	34	86.4
LigE-Y23F_fwd	TATTTGGTACGCCAAACAAACGGGCTAATGGTACAAC	37	77.8
LigE-Y23F_rev	GTTGTACCATTAGCCCGTTTGTTTGGCGTACCAAATA	37	77.8
LigE-Y122F_fwd	GGCAGGCTCAGATCATGAAAATCCAGGATATAACAAC	37	76.9
LigE-Y122F_rev	GTTGTTATATCCTGGATTTTCATGATCTGAGCCTGCC	37	76.9
LigE-R138A_fwd	TCCCAGAAACCACTGTTCAGCGCTCCAACGAACATAATC	39	82.7
LigE-R138A_rev	GATTATGTTCGTTGGAGCGCTGAACAGTGGTTTCTGGGA	39	82.7
LigF-F7A_fwd	CTATTTGCACCCGGACCAGCGCTATACAGTTTCAGGGTC	39	82.3
LigF-F7A_rev	GACCCTGAAACTGTATAGCGCTGGTCCGGGTGCAAATAG	39	82.3
LigF-N12A_fwd	GCGGTTTCAGGCTAGCTGCACCCGGACCAAAGC	33	85.3

LigF-N12A_rev	GCTTTGGTCCGGGTGCAGCTAGCCTGAAACCGC	33	85.3
LigF-C107A_fwd	TGGTGCTAACGCACCAAGCGAAATATTCATCAACCCATTTGG	42	84.2
LigF-C107A_rev	CCAAATGGGTTGATGAATATTTCGCTTGGTGCGTTAGCACCA	42	84.2
LigF-W108A_fwd	ATGGTGCTAACGCACGCACAGAAATATTCATCAACCCATTTGGT	44	83.6
LigF-W108A_rev	ACCAAATGGGTTGATGAATATTTCTGTGCGTGCGTTAGCACCAT	44	83.6
LigF-V110A_fwd	ATATTTCTGTTGGTGCGCTAGCACCATTGGTTGGG	35	80.3
LigF-V110A_rev	CCCAACCAATGGTGCTAGCGCACCAACAGAAATAT	35	80.3
LigF-S111A_fwd	CCAAATGCCCAACCAATGGTAGCAACGCACCAACAGAAATATTCATCAAC	50	85.9
LigF-S111A_rev	GTTGATGAATATTTCTGTTGGTGCGTTGCTACCATTGGTTGGGCATTTGG	50	85.9
LigF-W115A_fwd	ATTGCTTTAATACCAAATGCAGCACCAATGGTGCTAACGACACAACAG	48	84.5
LigF-W115A_rev	CTGTTGGTGCGTTAGCACCATTGGTGCTGCATTTGGTATTAAAGCAAT	48	84.5
LigF-I119A_fwd	CATTTTCTGGGCAATTGCTTTAGCACCAAATGCCCAACCAATGGT	45	85.6
LigF-I119A_rev	ACCATTGGTTGGGCATTTGGTGCTAAAGCAATTGCCCAGAAAATG	45	85.6
LigF-I122A_fwd	GCCATTTCAGCTGCTGTTCAGCGATCGGAACGTTTTTATTG	41	83.6
LigF-I122A_rev	CAATAAAAACGTTCCGATCGCTGAACAGCAGCTGAAATGGC	41	83.6
LigF-P142A_fwd	GCTCATTTTCTGGGCAGCTGCTTTAATACCAAATGCCCAACCAATGG	47	85.6
LigF-P142A_rev	CCATTGGTTGGGCATTTGGTATTAAAGCAGCTGCCCAGAAAATGAGC	47	85.6
LigF-Q144A_fwd	CGACGCCATTTCAGCTGAGCTTCCGGGATCGGAACGTTT	39	87.8
LigF-Q144A_rev	AAACGTTCCGATCCCGGAAGCTCAGCTGAAAYGGCGTCG	39	87.8
LigF-K147A_fwd	TACGTGCACGACGCCAAGCCAGCTGCTGTTCCGGG	35	90.0
LigF-K147A_rev	CCCGGAACAGCAGCTGGCTTGGCGTCGTGCACGTA	35	90.0
LigF-W148A_fwd	TGCACGACGCGCTTTCAGCTGCTGTTCCGGG	31	88.1
LigF-W148A_rev	CCCGGAACAGCAGCTGAAAGCGCGTCGTGCA	31	88.1
LigF-I199A_fwd	GACGCTGCAGACCATTTGCAGCGGCAAAATTACAAATATCGGC	43	85.9
LigF-I199A_rev	GCCGATATTTGTAATTTTGCCGCTGCAAATGGTCTGCAGCGTC	43	85.9
LigF-Q39A_fwd	AATCTTTTTGAACCAATCGCTATGAGCCTCAAATTTGCTCGGATCCACAAA	51	83.8
LigF-Q39A_rev	TTTGTGGATCCGAGCAAATTTGAGGCTCATAGCGATTGGTTCAAAAAGATT	51	83.8
LigF-W115R_fwd	TTAATACCAAATGCCCTACCAATGGTGCTAACGCAC	36	77.7
LigF-W115R_rev	GTGCGTTAGCACCATTGGTAGGGCATTTGGTATTAA	36	77.7
LigF-W115N_fwd	ATTGCTTTAATACCAAATGCGTTACCAATGGTGCTAACGCACCAACAG	48	82.7
LigF-W115N_rev	CTGTTGGTGCGTTAGCACCATTGGTAACGCATTTGGTATTAAAGCAAT	48	82.7
LigF-W115D_fwd	ATTGCTTTAATACCAAATGCGTCACCAATGGTGCTAACGCACCAACAG	48	84.5
LigF-W115D_rev	CTGTTGGTGCGTTAGCACCATTGGTGACGCATTTGGTATTAAAGCAAT	48	84.5
LigF-W115C_fwd	TTTAATACCAAATGCACAACCAATGGTGCTAACGCACC	38	79.6
LigF-W115C_rev	GGTGCGTTAGCACCATTGGTTGTGCATTTGGTATTAAA	38	79.6
LigF-W115Q_fwd	GCTTTAATACCAAATGCCTGACCAATGGTGCTAACGCACCAA	42	82.6
LigF-W115Q_rev	TTGGTGCGTTAGCACCATTGGTCAGGCATTTGGTATTAAAGC	42	82.6
LigF-W115E_fwd	GCTTTAATACCAAATGCCTCACCAATGGTGCTAACGCACCAA	42	82.6
LigF-W115E_rev	TTGGTGCGTTAGCACCATTGGTGAGGCATTTGGTATTAAAGC	42	82.6
LigF-W115G_fwd	TTAATACCAAATGCCCCACCAATGGTGCTAACGCAC	36	80.7
LigF-W115G_rev	GTGCGTTAGCACCATTGGTGGGGCATTTGGTATTAA	36	80.7
LigF-W115H_fwd	AATTGCTTTAATACCAAATGCATGACCAATGGTGCTAACGCACCAACAGA	51	78.5
LigF-W115H_rev	TCTGTTGGTGCGTTAGCACCATTGGTCATGCATTTGGTATTAAAGCAATT	51	78.5
LigF-W115I_fwd	AATTGCTTTAATACCAAATGCTATACCAATGGTGCTAACGCACCAACAGA	51	82.0
LigF-W115I_rev	TCTGTTGGTGCGTTAGCACCATTGGTATAGCATTTGGTATTAAAGCAATT	51	82.0
LigF-W115L_fwd	GCTTTAATACCAAATGCCAGACCAATGGTGCTAACGCACCAA	42	82.6
LigF-W115L_rev	TTGGTGCGTTAGCACCATTGGTCTGGCATTTGGTATTAAAGC	42	82.6
LigF-W115K_fwd	GCTTTAATACCAAATGCCTTACCAATGGTGCTAACGCACCAA	42	80.7
LigF-W115K_rev	TTGGTGCGTTAGCACCATTGGTAAGGCATTTGGTATTAAAGC	42	80.7
LigF-W115M_fwd	GCTTTAATACCAAATGCCATACCAATGGTGCTAACGCACCAA	42	81.2
LigF-W115M_rev	TTGGTGCGTTAGCACCATTGGTATGGCATTTGGTATTAAAGC	42	81.2
LigF-W115F_fwd	TGCTTTAATACCAAATGCAAAACCAATGGTGCTAACGCACCAAC	44	82.2
LigF-W115F_rev	GTTGGTGCGTTAGCACCATTGGTTTTGCATTTGGTATTAAAGCA	44	82.2
LigF-W115P_fwd	CTTTAATACCAAATGCCGGACCAATGGTGCTAACGCACCA	40	83.1
LigF-W115P_rev	TGGTGCGTTAGCACCATTGGTCCGGCATTTGGTATTAAAG	40	83.1
LigF-W115S_fwd	GCTTTAATACCAAATGCACTACCAATGGTGCTAACGCACCAA	42	80.5

LigF-W115S_rev	TTGGTGCGTTAGCACCATTGGTAGTGCATTTGGTATTAAAGC	42	80.5
LigF-W115T_fwd	TGGTGCGTTAGCACCATTGGTACGGCATTTGGTATTAAAG	40	80.9
LigF-W115T_rev	CTTTAATACCAAATGCCGTACCAATGGTGCTAACGCACCA	40	80.9
LigF-W115Y_fwd	GCTTTAATACCAAATGCGTAACCAATGGTGCTAACGCACCAA	42	81.2
LigF-W115Y_rev	TTGGTGCGTTAGCACCATTGGTTACGCATTTGGTATTAAAGC	42	81.2
LigF-W115V_fwd	GCTTTAATACCAAATGCCACACCAATGGTGCTAACGCACCAA	42	83.1
LigF-W115V_rev	TTGGTGCGTTAGCACCATTGGTGTGGCATTTGGTATTAAAGC	42	83.1
LigF-A11F_fwd	GCCAGCGGTTTCAGGCTATTAAAACCCGGACCAAAGCTATACA	43	78.9
LigF-A11F_rev	TGTATAGCTTTGGTCCGGGTTTTAATAGCCTGAAACCGCTGGC	43	78.9
LigF-S13A_fwd	GTTGCCAGCGGTTTCAGTGCATTTGCACCCGGACCAAAGC	40	89.4
LigF-S13A_rev	GCTTTGGTCCGGGTGCAAATGCACTGAAACCGCTGGCAAC	40	89.4
LigF-N12A, S13A_fwd	GTTGCCAGCGGTTTCAGGGCAGCTGCACCCGGACCAAAGCTAT	43	90.3
LigF-N12A, S13A_rev	ATACGTTTGGTCCGGGTGCAGCTGCCCTGAAACCGCTGGCAAC	43	90.3
Site saturation oligonucleotides			
LigF-P142X_fwd (NNS)	GACGCCATTTCAGCTGCTGTTCSNNGATCGGAACGTTTTTATTGATG	47	85.2
LigF-P142X_rev (NNS)	CATCAATAAAAACGTTCCGATCNNSGAACAGCAGCTGAAATGGCGTC	47	85.2
LigF-S13X_fwd (NNS)	GTTGCCAGCGGTTTCAGSNNATTTGCACCCGGACCAAAGC	40	78.9
LigF-S13X_rev (NNS)	GCTTTGGTCCGGGTGCAAATNNSCTGAAACCGCTGGCAAC	40	78.9
LigF-N12X_fwd (NNS)	CCAGCGGTTTCAGGCTSNNTGCACCCGGACCAAAGCTAT	39	78.6
LigF-N12X_rev (NNS)	ATAGCTTTGGTCCGGGTGCANNSAGCCTGAAACCGCTGG	39	78.6
LigE-S21X_fwd (NNS)	CGCCAAACATACGGSNNAATGGTACAACCGCTTTCCAGCTGCA	43	78.9
LigE-S21X_rev (NNS)	TGCAGCTGGAAAGCGGTTGTACCATTNNSCCGTATGTTTGGCG	43	78.9

2.1.4 Commercial enzymes, enzyme kits and dyes.

Tab. 2.7: Commercial enzymes, enzyme kits and dyes used in this study.

Enzyme/kit	Source	Supplier
NcoI-HF	*Nocardia carolina*	New England Biolabs (Ipswich, USA)
HindIII-HF	*Haemophilus influenzae* Rd	New England Biolabs (Ipswich, USA)
EcoRI	*Escherichia coli* RY13	New England Biolabs (Ipswich, USA)
NdeI-HF	*Neisseria denitrificans*	New England Biolabs (Ipswich, USA)
XhoI-HF	*Xanthomonas holcicola*	New England Biolabs (Ipswich, USA)
EcoRV	plasmid J62 pLG74	New England Biolabs (Ipswich, USA)
DpnI	*Diplococcus pneumoniae* G41	New England Biolabs (Ipswich, USA)
CutSmart buffer		New England Biolabs (Ipswich, USA)
T4 DNA Ligase	Bacteriophage T4	New England Biolabs (Ipswich, USA)
T4 DNA Ligase buffer		New England Biolabs (Ipswich, USA)
DreamTaq DNA Polymerase	*Thermus aquaticus*	Thermo Fisher Scientific (Waltham, USA)
PfuTurbo DNA polymerase	*Pyrococcus furiosus*	Agilent Technologies, (Santa Clara, USA)
Phusion© polymerase	*Pyrococcus furiosus*	New England Biolabs (Ipswich, USA)
PfuUltra II Hotstart Master Mix (2x)	*Pyrococcus furiosus*	Agilent Technologies, (Santa Clara, USA)
Green Go Taq® Flexi Buffer (10x)		Promega (Madison, USA)
Midori Green Direct		NIPPON Genetics Europe (Düren, DE)
HDGreen™ DNA Stain		INTAS (Göttingen, DE)

2.1.5 β-etherases and glutathione lyases

Tab. 2.8: β-etherases and glutathione lyases used in the presented study.

Enzyme	Source	NCBI number
LigE	*Sphingobium paucimobilis SYK-6*	WP_014075192.1
LigE-NS	*Novosphingobium* sp. PP1Y	WP_013832481.1
LigE-NA	*Novosphingobium aromaticivorans* DSM1244	WP_011446047.1
LigP	*Sphingobium paucimobilis SYK-6*	WP_014077574.1
LigF	*Sphingobium paucimobilis SYK-6*	WP_014075191.1
LigF-NS	*Novosphingobium sp. PP1Y*	WP_013832480.1
LigF-NA	*Novosphingobium aromaticivorans DSM1244*	WP_041551020.1
LigG	*Sphingobium paucimobilis SYK-6*	WP_041392591.1
LigG-NS	*Novosphingobium* sp. PP1Y	WP_041558818.1
LigG-TD	*Thiobacillus denitrificans ATCC 25259*	WP_011311562.1

2.1.6 Growth media and buffer solutions

Tab. 2.9: Growth media and buffer solutions used in this study with their components.

Growth medium/buffer solution	Component	Preparation
Lysogeny broth (LB media)	- 10 g/L tryptone - 5 g/L yeast extract - 10 g/L sodium chloride	-mixed with MilliQ water up to desired volume, autoclaved
Lysogen broth agar (LB agar)	- 10 g/L tryptone - 5 g/L yeast extract - 10 g/L sodium chloride - 15 g/L agar-agar	-mixed with MilliQ water up to desired volume, autoclaved, (possibly added antibiotic) poured to the Petri dish
Terrific broth (TB media)	- 12 g/L tryptone - 24 g/L yeast extract - 5 g/L glycerol	-mixed with MilliQ water up to 90 % v/v volume, autoclaved, filled up with 10x TB salts
Terrific broth salts – 10x concentrated (TB salts)	- 23.1 g/L KH_2PO_4 - 125.4 g/L K_2HPO_4	-mixed with MilliQ water up to desired volume, autoclaved
Super optimal broth with catabolite repression (SOC media)	- 20 g/L tryptone - 5 g/L yeast extract - 0.5 g/L sodium chloride - 0.186 g/L potassium chloride - 10 mM magnesium chloride - 10 mM magnesium sulfate - 20 mM glucose	-solution containing tryptone, yeast extract, sodium chloride, potassium chloride mixed with MilliQ water up to desired volume, autoclaved -solution of 2 M glucose (sterilized by filtration) -solutions of 2 M magnesium sulfate and 2 M magnesium chloride, autoclaved -add glucose (1:100), magnesium sulfate (1:200) and magnesium chloride (1:200) to the main broth
IMAC* binding buffer	- 20 mM KH_2PO_4 - 500 mM sodium chloride - 20 mM imidazole	-mixed with MilliQ water up to desired volume, pH adjusted by HCl to 7.4, filtered, degassed
IMAC elution buffer	- 20 mM KH_2PO_4 - 500 mM sodium chloride - 500 mM imidazole	-mixed with MilliQ water up to desired volume, pH adjusted by HCl to 7.4, filtered, degassed
IMAC pre-elution buffer **	- 20 mM KH_2PO_4 - 500 mM sodium chloride - 125 mM imidazole	-mixed with MilliQ water up to desired volume, pH adjusted by HCl to 7.4, filtered, degassed
Protein storage buffer	- 20 mM TRIS-base - 20 % v/v glycerol	-mixed with MilliQ water up to desired volume, pH adjusted by HCl to 7.5, autoclaved
Tris-acetate-EDTA buffer (TAE buffer)	- 4.84 g/L TRIS-base - 1.142 mL/L acetic/glacial acid - 2 mL/L of 0.5 mM EDTA	-mixed with MilliQ water up to desired volume

SDS PAGE-running buffer	- 3.03 g/L TRIS-base - 14.44 g/L glycine - 1.0 g/L sodium dodecyl sulfate (SDS)	-mixed with MilliQ water up to desired volume
SDS PAGE-sample buffer 4x concentrated	- 80 mg/L SDS - 40 % v/v glycerol - 20 % v/v mercaptoethanol - 4 mg/L bromophenol blue - 100 mM TRIS-HCl, pH 6.8	-mixed SDS with TRIS-HCl, added rest of the components
SDS PAGE-gel staining solution	- 30 % v/v ethanol - 10 % v/v acetic/glacial acid - 2.5 g/L Coomassie Brilliant Blue R-250®	-mixed with MilliQ water up to desired volume
SDS PAGE-gel destaining solution	- 30 % v/v ethanol - 10 % v/v acetic/glacial acid	-mixed with MilliQ water up to desired volume
RF 1 buffer	- 30 mM KCH_3CO_2 - 50 mM $MnCl_2$ - 100 mM RbCl - 10 mM $CaCl_2$ - 15 % v/v glycerol, pH 5.8	-mixed with MilliQ water up to desired volume -pH adjusted with acetic/glacial acid -sterilized by filtration
RF 2 buffer	- 10 mM MOPS - 75 mM $CaCl_2$ - 10 mM $RbCl_2$ - 15 % v/v glycerol, pH 6.8	-mixed with MilliQ water up to desired volume -pH adjusted with NaOH -sterilized by filtration

* IMAC= immobilized metal affinity chromatography

** used in gravity flow purification

2.2 Methods

2.2.1 Synthesis of model substrates

Several substrates used for the analysis and description of the biocatalysts were synthesized after modified protocols form Picart *et al*. [110]. The three-step synthesis included bromination of an acetophenone derivate, followed by S_N2 reaction with phenol compound, forming the keto-ether bond, and finished by the hydroxy-methylation of the keto-ether compound using paraformaldehyde. General scheme of synthesis is shown in **Fig. 2.1** and the shortcuts, abbreviations and the full names of the compounds are listed in **Tab. 2.10**. NMR analyses of synthesized compounds are shown in chapter 7.5.

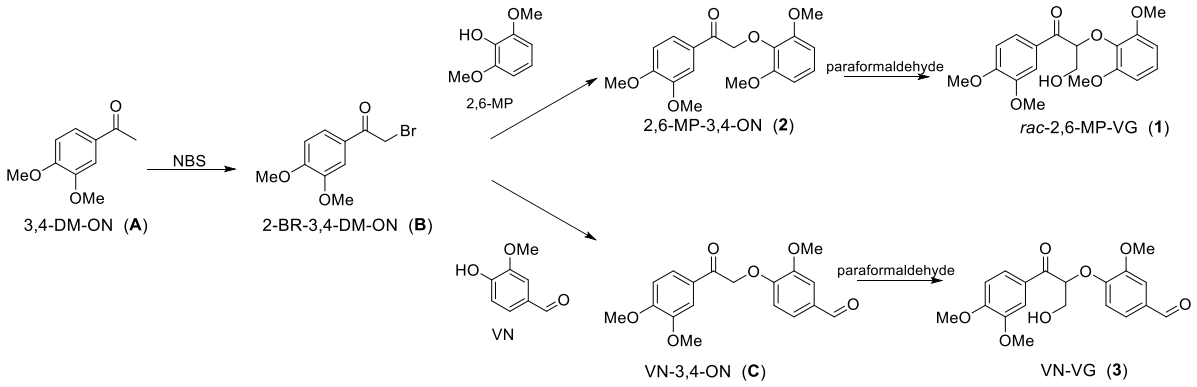


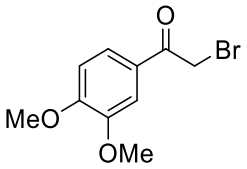
Fig. 2.1: General scheme of substrates synthesis.

Tab. 2.10: Names, abbreviations and shortcuts of substrates and their precursors synthesized.

Compound shortcut	Compound abbreviation	Compound full name
A	3,4-DM-ON	3,4-dimethoxyacetophenon
B	2-BR-3,4-DM-ON	2-bromo-1-(3,4-dimethoxyphenyl)-ethan-1-one
C	VN-3,4-ON	4-(2-(3,4-dimethoxyphenyl)-2-oxoethoxy)-3-methoxybenzaldehyde
3	VN-VG	β-vanillinyl-α-veratrylglycerone
2	2,6-MP-3,4-ON	β-(2,6-dimethoxyphenoxy)-α-(3,4-dimethoxyphenyl)-ethan-1-on
1	*rac*-2,6-MP-VG	β-(2,6-dimethoxyphenoxy)-α-veratrylglycerone

2.2.1.1 Synthesis of 2-bromo-1-(3,4-dimethoxyphenyl)-ethan-1-one (B)

A dry and argon-flushed three-necked flask equipped with a magnetic stirrer, a reflux condenser, an argon inlet and a septum were charged with 5.0 g of **A** (1.0 eq., 28 mmol), 8.0 g of *p*-toluenesulfonic acid monohydrate (1.5 eq., 42 mmol) and 400 mL of dry acetonitrile (dried overnight with activated molecular sieve with 3 Å pore size). After all the substrates were dissolved, a solution of 6.0 g of N-bromosuccinimide (NBS) (1.2 eq., 33.6 mmol) in 100 mL of dry acetonitrile was added. The reaction mixture was stirred for 2 h at 100 °C and upon completion cooled to room temperature. The solvent was evaporated under reduced pressure, the residue dissolved in dichloromethane and washed with distilled water. The organic phase was dried over $MgSO_4$, filtered and evaporated under reduced pressure. The product was purified by column chromatography (dichloromethane/ethyl acetate, 100:1) yielding the product **B** as white solid with yield of 70 % (5.0 g).



^{1}H NMR (400 MHz, CCl_3D): δ [ppm] = 7.60 (dd, ^{3}J = 8.4 HZ, ^{5}J = 2.0 Hz, 1H), 7.54 (d, ^{5}J = 2.0 Hz, 1H), 6.92 (d, ^{3}J = 8.2 Hz, 1H), 4.40 (s, 2H), 3.96 (s, 3H), 3.95 (s, 3H)

^{13}C NMR (100 MHz, CCl_3D): δ [ppm] = 190.2, 154.3, 149.6, 127.0, 124.0, 111.1, 110.1, 56.3, 56.1, 30.6

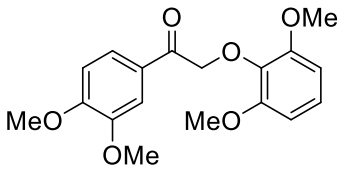
2.2.1.2 Synthesis of β-keto ethers (2, C)

5.0 g of **B** (1.0 eq.0, 19.3 mmol), 3.0 g of anhydrous Na_2CO_3 (1.5 eq., 29.0 mmol) and the corresponding phenol compound (for synthesis of **2**; 2,6-methoxyphenol, for synthesis of **C**; vanillin) (1.5 eq., 29 mmol) were dissolved in 50 mL of acetone. The reaction mixture was stirred at room temperature for 72 hours. After filtration and washing with dichloromethane, the solution was evaporated under reduced pressure. The residue

was re-dissolved in dichloromethane and washed with distilled water and brine. The organic phase was dried over $MgSO_4$, filtered and evaporated under reduced pressure. The product was purified by column chromatography.

Synthesis of 2

The crude product was purified by column chromatography with gradient from pure dichloromethane to dichloromethane:ethyl acetate of 4:1. The product **2** was isolated with a yield of 59.4 % (3.8 g).



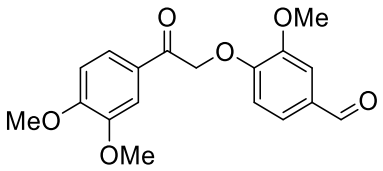
^{1}H NMR (600 MHz, CCl_3D): δ [ppm] = 7.71 (dd, ^3J = 8.5 Hz, ^5J = 2.0 Hz, 1H), 7.51 (d, ^5J = 2.0 Hz, 1H), 7.05 (d, ^3J = 8.2 Hz, 1H), 7.01 (t, 4.40 ^3J = 8.5 Hz, 1H), 6.69 (d, ^3J = 8.4 Hz, 2H), 5.07 (s, 2H), 3.86 (s, 3H), 3.85 (s, 3H), 3.74 (s, 6H)

^{13}C NMR (150 MHz, CCl_3D): δ [ppm] = 193.3, 153.4, 153.0, 148.0, 135.8, 127.6, 124.2, 122.8, 111.0, 110.3, 105.6, 74.6, 55.7, 55.6, 55.3

* the compound **2** was used as a precursor for the synthesis of the substrate **1**, but was also used as a substrate for the enzymes' analysis

Synthesis of C

The crude product was purified by column chromatography with gradient from pure dichloromethane to dichloromethane:ethyl acetate of 4:1. The product **C** was isolated with a yield of 87.2 % (5.6 g).



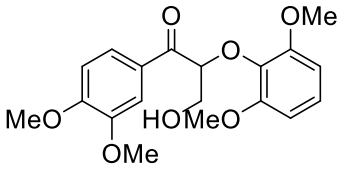
^{1}H NMR (600 MHz, DMSO): δ [ppm] = 9.83 (s, 1H), 7.71 (dd, ^3J = 8.4 Hz, ^5J = 2.0 Hz, 1H), 7.51 (d, ^5J = 2.0 Hz, 1H), 7.48 (dd, ^3J = 8.5 Hz, ^5J = 1.9 Hz, 1H), 7.42 (d, ^5J = 2.0 Hz, 1H), 7.13 (d, ^3J = 8.2 Hz, 1H), 7.05 (d, ^3J = 8.5 Hz, 1H), 5.69 (s, 2H), 3.86 (s, 3H), 3.85 (s, 3H), 3.80 (s, 3H)

^{13}C NMR (150 MHz, DMSO): δ [ppm] = 192.3, 191.5, 153.4, 153.0, 149.0,148.9, 130.1, 127.2, 125.2, 122.8, 112.5, 110.9, 110.3, 110.0, 70.6, 55.7, 55.6, 55.3

2.2.1.3 Hydroxy methylation of β-keto-ethers (1, 3)

Synthesis of 1

3.0 g of **2** (1.0 eq., 9 mmol), 1.9 mL of paraformaldehyde (10.0 eq., 90 mmol), and 9.54 g of anhydrous Na_2CO_3 (1.0 eq.) were dissolved in 30 mL of dry dichloromethane. The mixture was stirred at room temperature for 20 h. When the reaction was over, distilled water was added to the reaction and the aqueous phase was extracted with dichloromethane. The organic phase was washed with brine. The combined organic layers were dried over $MgSO_4$, filtered and evaporated under reduced pressure. The product was purified by column chromatography (dichloromethane:ethyl acetate, 100:1) yielding the product **1** as slightly brown oil with yield of 76.1 % (2.5 g).

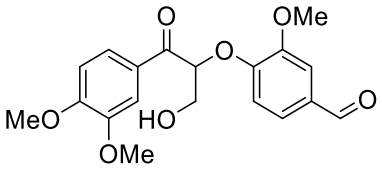


^{1}H NMR (600 MHz, DMSO): δ [ppm] = 7.72 (dd, ^3J = 8.5 Hz, ^5J = 2.1 Hz, 1H), 7.51 (d, ^3J = 2.1 Hz, 1H), 7.07 (d, ^3J = 8.5 Hz, 1H), 6.98 (t, ^3J = 8.3 Hz, 1H), 6.65 (d, ^3J = 8.3 Hz, 2H), 5.20 (m, 1H), 4.70 (m, 1H), 3.85 (s, 3H), 3.80 (s, 3H), 3.77 (m, 2H), 3.64 (s, 3H)

^{13}C NMR (150 MHz, DMSO): δ [ppm] = 195.4, 153.1, 152.5, 148.5, 136.0, 128.9, 123.7, 123.5, 110.8, 105.6, 83.5, 62.2, 55.86, 55.8, 55.5

Synthesis of 3

3.0 g of **C** (1.0 eq., 9 mmol), 1.9 mL of paraformaldehyde (10.0 eq., 90 mmol), and 4.77 g of anhydrous Na_2CO_3 (0.5 eq.) were dissolved in 30 mL of isopropyl alcohol. The mixture was stirred at 45˚C for 20 h. When the reaction was over, distilled water was added to the reaction and the aqueous phase was extracted with dichloromethane. The organic phase was washed with brine. The combined organic layers were dried over $MgSO_4$, filtered and evaporated under reduced pressure. The crude product was repeatedly purified by column chromatography with gradient from pure dichloromethane to dichloromethane:ethyl acetate of 1:1 yielding the product **3** as slightly yellow-white solid with yield of 62.3 % (2.0 g).



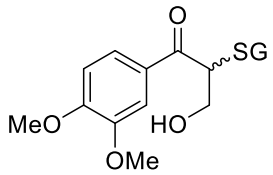
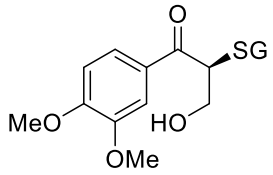
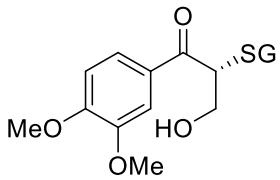
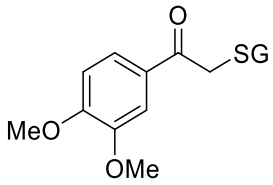
^{1}H NMR (600 MHz, DMSO): δ [ppm] = 9.81 (s, 1H), 7.81 (dd, ^{3}J = 8.4 Hz, ^{5}J = 2.0 Hz, 1H), 7.52 (d, ^{5}J = 2.0 Hz, 1H), 7.44 (m, ^{5}J = 1.9 Hz, 1H), 7.42 (m, 1H), 7.13 (d, ^{3}J = 8.5 Hz, 1H), 6.90 (d, ^{3}J = 8.8 Hz, 1H), 5.95 (m, 1H), 5.28 (t, ^{3}J = 5.7 Hz, 1H), 3.95 (m, 2H), 3.85 (s, 3H), 3.84 (s, 3H), 3.81 (s, 3H)
^{13}C NMR (150 MHz, DMSO): δ [ppm] = 194.3, 191.5, 153.7, 152.4, 149.0,148.9, 130.1, 127.2, 125.2, 123.6, 113.0, 110.9, 110.8, 110.2, 81.2, 63.0, 55.7, 55.6, 55.0

2.2.1.4 Enzymatic synthesis of glutathione adducts

In order to prepare the model substrates for glutathione lyases, the enzymatical reactions were performed. Racemic model substrates **11** and **12** were enzymatically synthesized in reactions where compound **1** or compound **2** were used as substrates. Reactions were performed in 5 mL volume and consisted of 0.2 mM of corresponding substrate (final DMSO concentration 5 % v/v), 1.0 mM of GSH, 100 mM glycine buffer pH 9.5 and the 50 µg of enzymes LigE and LigF-NS. After the completion of the enzymatic reaction, the volume of the reaction was mixed with 5 mL of water and 5 mL of ethyl acetate (EtOAc) forming two immiscible solvent phases system. The mixture was transformed into the extraction funnel and the organic phase was separated and the solvent was evaporated under reduced pressure.

For preparation of the (*R*)-**11** and (*S*)-**11** substrates, the reaction was modified and 0.4 mM of **1** was used as well as only LigE (for preparation of (*S*)-**11**) or only LigF-NS (for preparation of (*R*)-**11**) were employed in the reactions. In these cases, after the 50 % of the substrate **1** was converted to the desired enantiomeric product, the extraction step was omitted and the received mixture was used as a starting point for the characterization of glutathione lyases (fresh GSH had to be added with the new enzyme). All of the substrates prepared for the characterization of glutathione lyases with their structures, shortcuts, abbreviations and full names are summarized in the Tab. **2.11**.

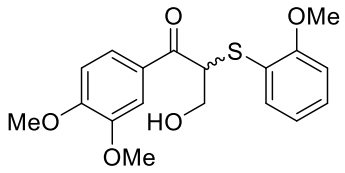
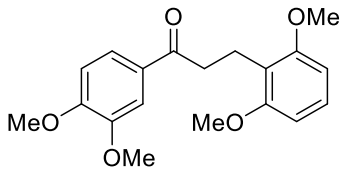
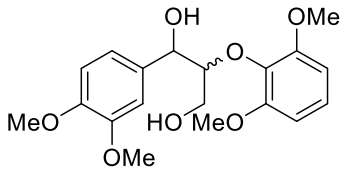
Tab. 2.11: Structures, names, abbreviations and shortcuts of substrates enzymatically synthesized for the characterization of glutathione lyases.

Compound structure	Shortcut	Abbreviation	Compound full name
	**11**	GS-VG	β-glutathionyl-α-veratrylglycerone
	(*R*)-**11**	(*R*)-GS-VG	(*R*)-β-glutathionyl-α-veratrylglycerone
	(*S*)-**11**	(*S*)-GS-VG	(*S*)-β-glutathionyl-α-veratrylglycerone
	**12**	GS-3,4-ON	β-glutathionyl-α--(3,4-dimethoxyphenyl)-ethan-1-on

2.2.1.5 Other substrates used in this thesis

Several other dimeric lignin-like model substrates were used in this thesis. Structures of these together with their shortcuts, abbreviations and full names are listed in **Tab. 2.12**. These substrates were synthesized by Dr. Hauke Voß and were kindly provided for the presented research. The protocols for synthesis of these molecules are precisely described in his doctoral thesis [269].

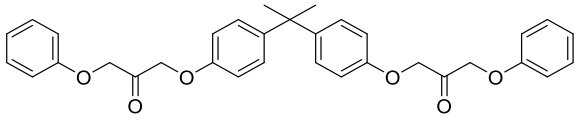
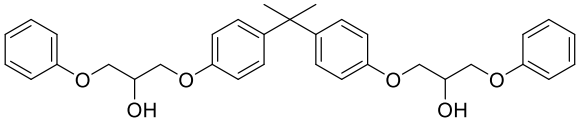
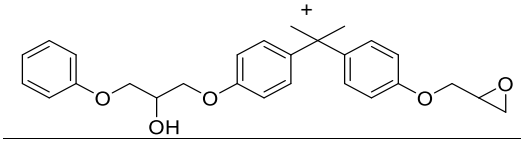
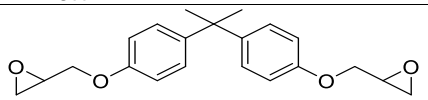
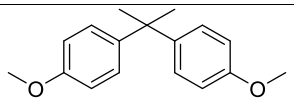
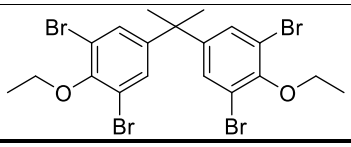
Tab. 2.12: Structures, names, abbreviations and shortcuts of substrates synthesized by Dr. Voß.

Compound structure	Shortcut	Abbreviation	Compound full name
	**4**	*rac*-MTP-VG	β-(2-methoxytiophenoxy)--α-veratrylglycerone
	**5**	2,6-MPL-3,4-ON	β-(2,6-dimethoxyphenyl)--α-(3,4-dimethoxyphenyl)-ethan-1-on
	**6**	*rac*-2,6-MP-VGL	β -(2,6-dimethoxyphenoxy)--α-veratrylglycerol

2.2.1.6 Non lignin-related model substrates

Non-lignin model substrates containing aryl ether bonds, which were used in this thesis, were provided by Prof. Dr. Dieter E. Kaufmann from the Institute of Organic Chemistry at TU Clausthal, Germany. These manmade compounds together with their names and shortcuts are shown in **Tab. 2.13**.

Tab. 2.13: Structures, names, and shortcuts of non-lignin-based substrates tested in this thesis.

Non-lignin substrate structure	Shortcut	Compound full name
	**30**	3,3'-((propane-2,2-diylbis (4,1-phenylene))bis(oxy))bis (1-phenoxypropan-2-one)
 + 	**31**	3,3'-((propane-2,2-diylbis (4,1-phenylene))bis(oxy))bis (1-phenoxypropan-2-ol) + 1-(4-(2-(4-(oxiran-2-ylmethoxy) phenyl)propan-2-yl)phenoxy)- 3-phenoxypropan-2-ol
	**32**	2,2'-(((propane-2,2-diyl bis(4,1-phenylene))bis(oxy))bis (methylene))bis(oxirane)
	**33**	4,4'-(propane-2,2-diyl)bis (methoxybenzene)
	**34**	5,5'-(propane-2,2-diyl)bis (1,3-dibromo-2-ethoxybenzene)

2.2.2 Molecular and microbiological methods

2.2.2.1 Gene synthesis

All of the genes used in presented work were synthesized by Eurofins (Ebersberg, Germany) with codon optimized sequences for heterologous expression in *E. coli* strains. The genes contained restriction enzymes 5'-NdeI and 3'-HindIII allowing simple restriction and subsequent subcloning into the pET-28a(+) plasmid.

2.2.2.2 Restriction digest and ligation

Restriction enzymes or restriction endonucleases type II recognize and bind to specific sequences of DNA. Each restriction enzyme recognizes one or a few restriction sites

and cuts a double-stranded DNA molecule in a tidy, predictable pattern. The DNA can be cut through both strands directly (blunt end) or with 5' or 3' overhangs (sticky ends). Using two different restriction enzymes to digest the target vector and desired DNA insert allows further specific ligations and creations of the vectors carrying requested DNA genes mentioned in chapter 2.1.3.1. Ligation reaction in molecular biology refers to the joining of the two or more DNA fragments through the formation of covalent, phosphodiester bond. The reaction is catalyzed using ATP-dependent T4-DNA ligase that connects complementary sticky ends of the DNA fragments (vector and insert) created in the digestion reaction.

Subcloning of β-etherases LigE and LigF in vector pETDuet-1

LigE gene located on the pET-28a(+) vector was amplified by PCR using forward primer pET28a_NdeI_to_NcoI_fwd to introduce a NcoI restriction site and reverse primer pET28a_universal_rev as described in chapter 2.2.2.4 (Gene of interest amplification). The resulting *ligE* gene was purified using NucleoSpin® Gel and PCR Clean-up Kit following the manufacturer's protocol. For the digestion reaction, pETDuet-1 vector (1000 ng) was combined with amplified *ligE* gene in molecular ratio 1:5 (vector:insert). To the reaction, three restriction enzymes; NcoI, HindIII and EcoRI, were added in 1 µL volume (EcoRI enzyme was added as an extra restriction enzyme to digest the undesired part of the vector to prevent self ligation in the further steps). Afterwards, 5 µL of CutSmart buffer were added and the reaction was filled up to 50 µL with MilliQ water. Restriction enzymes digestion reaction was performed for 2 hours at 37˚C. Afterwards, the enzymes of the reaction were inactivated by exposing to the 85˚C for 20 min. From the received reaction, 6 µL of the product were analyzed on 0.8 % agarose gel as described in chapter 2.2.2.5 (Agarose gel electrophoresis). Subsequently, with the remaining digestion product (44 µL) the ligation reaction was performed. To the mixture, 5 µL of T4 DNA Ligase buffer and 1 µL of T4 DNA Ligase were added. The ligation reaction was carried out for 2 hours at 20˚C and was followed by enzyme inactivation at 65˚C for 30 min. After transformation of chemically competent *E. coli* DH5α, the successful insertion of the *ligE* in the MSC-1 of pETDuet-1 was verified by colony PCR using the protocol described in chapter 2.2.2.4 (Colony PCR) and sequencing.

Further, *ligF-NA* gene was amplified from pET-28a(+) by PCR using forward primer pET28a_universal_fwd and reverse primer pET28a_universal_rev as described in chapter 2.2.2.4 (Gene of interest amplification). The resulting *ligF* gene was purified using NucleoSpin® Gel and PCR Clean-up Kit following the manufacturer's protocol. For the digestion reaction, pETDuet-1_*ligE* vector (1000 ng) was combined with amplified *ligF* gene in molecular ratio 1:5 (vector:insert). To the reaction, three restriction enzymes;

NdeI, XhoI and EcoRV, were added in 1 µL volume (EcoRV enzyme was added as an extra restriction enzyme to digest the undesired part of the vector to prevent self ligation in the further steps). Afterwards, 5 µL of CutSmart buffer were added and the reaction was filled up to 50 µL with MilliQ water. Restriction enzymes digestion reaction was performed for 2 hours at 37˚C. Afterwards, the enzymes of the reaction were inactivated by exposing to the 85˚C for 20 min. From the received reaction, 6 µL of the product were analyzed on 0.8 % agarose gel as described in chapter 2.2.2.5 (Agarose gel electrophoresis). The remaining digestion product (44 µL) was used for the subsequent ligation reaction. To the mixture, 5 µL of T4 DNA Ligase buffer and 1 µL of T4 DNA Ligase were added. The ligation reaction was carried out for 2 hours at 20˚C and was followed by enzyme inactivation at 65˚C for 30 min. After transformation of chemically competent *E. coli* DH5α, the successful insertion of the *ligF* in the MSC-2 of pETDuet-1_*ligE* was verified by colony PCR using the protocol described in chapter 2.2.2.4 (Colony PCR) and sequencing. Finally, plasmid DNA was isolated with E.Z.N.A.® Plasmid Mini Kit I and digested using the endonucleases NcoI, HindIII, NdeI and XhoI to verify the presence of the two desired inserts. Additionally, an aliquot was sent for sequencing to Eurofins Genomics.

Cloning of glutathione lyase in vector pIT2-MCS

LigG-TD gene located on the pET-28a(+) vector was amplified by PCR using forward primer pET28a_universal_fwd and reverse primer pET28a_universal_rev as described in chapter 2.2.2.4 (Gene of interest amplification). The resulting *ligG-TD* gene was purified using NucleoSpin® Gel and PCR Clean-up Kit following the protocol. For the digestion reaction, pIT2 vector (1000ng) was combined with amplified *ligG-TD* gene in molecular ratio 1:5 (vector:insert). To the reaction, two restriction enzymes NcoI, and HindIII were added, each 1 µL volume. Afterwards, 5 µL of CutSmart buffer were added and the reaction was filled up to 50 µL with MilliQ water. Restriction enzymes digestion reaction was performed for 2 hours at 37˚C. Afterwards, the enzymes of the reaction were inactivated by exposing to the 85˚C for 20 min. From the received reaction, 6 µL of the product were analyzed on 0.8 % agarose gel as described in chapter 2.2.2.5 (Agarose gel electrophoresis). Subsequently, with the remaining digestion product (44 µL) the ligation reaction was performed. To the mixture, 5 µL of T4 DNA Ligase buffer and 1 µL of T4 DNA Ligase were added. The ligation reaction was carried out for 2 hours at 20˚C and was followed by enzyme inactivation at 65˚C for 30 min. After transformation of chemically competent *E. coli* DH5α, the successful insertion of the *ligG-TD* in the pIT2 was verified by colony PCR using the protocol described in chapter 2.2.2.4 (Colony PCR) and sequencing.

2.2.2.3 Preparation and transformation of chemically competent cells

The competence is an ability of the cell to alter its genetical information and uptake the foreign/extracellular DNA through transformation process. Some prokaryotic cells bear natural competence, while in the *E. coli* cells, the competence has to be achieved artificially. Chemically competent cells are transiently permeable to DNA, which is caused by several chemical treatments. In this work chemically competent cells were prepared from *E. coli* DH5α, *E. coli* BL21 (DE3), *E. coli* BL21 Gold (DE3) and *E. coli* C43 (DE3).

Preparation of chemically competent cells

For the preparation of chemically competent cells of the desired bacterium, 50 mL of LB medium was inoculated with an overnight culture to an optical density at 600 nm (OD_{600}) of 0.1. The culture was incubated at 37 °C at 250 rpm until an OD_{600} between 0.4 and 0.6 was reached. Afterwards, the cells were harvested by centrifugation (4000 rpm, 4 °C, and 30 min) and resuspended in 20 mL of ice-cold RF 1 buffer. Suspension was kept on ice for 30 min. Subsequently, the mixture was centrifuged and the resulting pellet was resuspended in 4 mL of ice-cold RF 2 buffer. Prepared suspension was aliquoted by 100 µL into the sterilized and precooled 1.5 mL tubes. Aliquots were shock frozen in liquid nitrogen and stored at -80 °C.

Transformation of chemically competent cells

Chemically competent *E. coli* cells were thaw on ice and 50 to 100 ng of plasmid DNA or ligation product was added to a cell aliquot. The cells were mixed with the DNA and afterwards incubated for 20 min on ice. Subsequently, the heat shock was performed for 90 s at 42°C. After the heat shock, the cell mixture was cooled down on ice for 5 min. Thereafter, 500 µL of pre-warmed SOC medium was added to the cell mixture. Cells were incubated for 1 h at 37 °C at 800 rpm for cell recovery and antibiotic resistance expression. At last, the transformed cells were plated on LB agar plates containing the appropriate selection marker (antibiotic, **Tab. 2.5**) and grown overnight at 37 °C.

2.2.2.4 Polymerase chain reaction (PCR)

The polymerase chain reaction is an extremely flexible, multi-purpose laboratory technique allowing the amplification of a selected DNA sequence. The oligonucleotides (primers) binding to the DNA template, define the start of the amplification. The process of amplification consists of three steps: denaturation (breaking the hydrogen bonds between complementary bases of the double-stranded DNA template, yielding two

single-stranded DNA molecules), annealing (binding of the oligonucleotides to each of the single-stranded DNA templates) and elongation (the DNA polymerase is synthesizing a new DNA strand, complementary to the DNA template strand, framed by the oligonucleotides in the 5'-to-3' direction). In this work, the PCR was applied to selectively modify the restriction site of the gene coding for LigE, to verify if the ligation was successful (colony PCR) and to introduce mutations in *ligE* and *ligF-NA* genes via site-directed mutagenesis.

Gene amplification

50-100 ng template DNA was added to a solution containing 1.0 U of Phusion© polymerase (1.0 µL), 20 % v/v of 5x Phusion HF buffer (10 µL), 200 µM of dNTPs (1 µL) and 3 % v/v of DMSO for microbiology (1.5 µL). The target DNA sequence was then selectively amplified using designed forward and reverse primers to a final concentration of 0.5 µM each. Finally, the volume was adjusted to 50 µL with MilliQ water. The gene of interest was amplified according to the temperature profile in **Tab. 2.14**.

Tab. 2.14: Temperature profile of the PCR reaction for the gene amplification.

Step	Temperature (˚C)	Time (s)	Number of cycles
1	98	120	1
2	98	20	
3	55	30	30
4	72	30 s per kb	
5	72	6000	1

DNA fragments of interest were separated via agarose gel electrophoresis and extracted from agarose gel using NucleoSpin® Gel and PCR Clean-up Kit prior to further experiments.

Colony PCR

Colony PCR is an effective, high-throughput method designed to quickly identify if the ligation reaction was successful and DNA fragment was incorporated in to the new vector backbone. After *E. coli* DH5α was transformed with the ligation product and the cells were grown on LB agar with adequate antibiotic, several single colonies were picked to be analyzed by colony PCR. An individual colony was picked with the sterile pipette tip and visible number of cells were transported to the sterile pre-cooled PCR tube (providing the DNA template for the PCR reaction). Remaining cells from the individual colony could be used for separate cultivation, if the colony is needed. Several single colonies were reprocessed. The PCR tubes with the cells were filled with 20 µL of the “in house” made PCR master mix (**Tab. 2.15**), mixed and the gene, framed by the used oligonucleotides, was amplified according to the temperature profile in **Tab. 2.16**. After

amplification, the mix was directly loaded on the agarose gel and the results directly visualized.

Tab. 2.15: Components of the "in house" made master mix used for colony PCR. Amounts of the shown components are necessary for the PCR of one colony. The master mix was prepared for x+1 reaction for x colonies, guarantying even composition in every reaction.

Component	Amount (µL)
Green Go Taq® Flexi Buffer (10x)	2.0
10 mM dNTP's	0.4
10 µM fwd_oligonucleotide	1.0
10 µM rev_oligonucleotide	1.0
DreamTaq DNA Polymerase	0.1
MilliQ water	15.5

Tab. 2.16: Temperature profile of the colony PCR reaction.

Step	Temperature (˚C)	Time (min)	Number of cycles
1	95	5	1
2	95	1	
3	55	0.5	30
4	72	1	
5	72	3	1

QuikChange ® PCR

The site-directed mutagenesis and site-saturation mutagenesis was performed using QuikChange ® PCR strategy. The desired mutations were introduced to the DNA using set of complementary oligonucleotides designed by QuikChange® Primer Design webtool from Agilent. These oligonucleotides contain set of altered codons, introducing the desired change to the DNA sequence. Components of the of the QuikChange® PCR reaction are listed in the **Tab. 2.17**. After all of the components were mixed, the amplification of the vector was performed according to the temperature profile in **Tab. 2.18**.

Tab. 2.17: Components of the reaction mixture for the QuikChange ® PCR reaction.

Component	Amount (µL)
PfuUltra II Hotstart Master Mix (2x)	25.0
10 µM fwd_QuikChange_oligonucleotide	1.0
10 µM rev_QuikChange_oligonucleotide	1.0
DNA template (100 ng/µL)	1.0
DMSO for microbiology	2.5
MilliQ water	19.5

Tab. 2.18: Temperature profile of the QuikChange ® PCR reaction.

Step	Temperature (˚C)	Time (s)	Number of cycles
1	95	120	1
2	95	20	
3	55	20	20
4	72	15 s per kb	
5	72	180	1

After the QuikChange® PCR reaction, 1 µL of *Dpn*I was added to the reaction to digest the DNA template vector. DpnI digestion was performed for 2 h at 37˚C. PCR product was purified by E.Z.N.A.® MicroElute DNA Clean Up Kit and analyzed.

2.2.2.5 Agarose gel electrophoresis

Agarose gel electrophoresis allows separation of the DNA fragments based on their size. The natural negative charge of the DNA caused by the phosphodiester bond in the backbone causes migration of the DNA towards the anode in the presence of electric field. The size separation is achieved by the agarose gel network hindering the larger molecules to a greater extent compared to the small molecules.

Agarose gel electrophoresis was generally used for plasmid visualization and vector insert separation after restriction enzyme's digestion. Agarose (0.8 or 1 % w/v) was mixes with TAE buffer and the mixture was heated up until dissolved. Hot agarose mixture was used to form an agarose gel. DNA samples were mixed with 6x sample buffer containing 1 % v/v of Midori Green Direct, and typically 10-20 µL were loaded on the gel together with 1 kb ladder (with Midori Green Direct) to reference the size of the DNA fragments. The separation based on molecular weight was performed using an electric current of 100 V. The TAE buffer was used as running buffer. The gel was visualized via GelPic LED Box.

Alternation was done for the colony PCR, where to the lukewarm agarose mixture 0.0025 % v/v of the HDGreen^{TD} DNA Stain was added. The 10 µL of the colony PCR reactions were loaded to the agarose gel together with 1 kb ladder (with Midori Green Direct) to reference the size of the DNA fragments. The separation based on molecular weight was performed using an electric current of 100 V. The TAE buffer was used as running buffer. The gel was visualized via GelPic LED Box.

2.2.2.6 Plasmid isolation

The plasmid DNA as a foreign DNA was amplified by an *E. coli* DH5α strain. In order to confirm the successful ligation of a DNA fragment in a vector backbone or the presence of a desired mutation within a gene, plasmid DNA had to be isolated from the transformed

E. coli DH5α strain. For plasmid DNA isolation, the E.Z.N.A.® Plasmid Mini Kit was used based on the manufacturer's instruction.

2.2.2.7 DNA sequencing

Isolated plasmid DNA was analyzed, and successful cloning and mutagenesis experiments were confirmed by sequencing at the company Eurofins Genomics (Ebersberg, Germany).

2.2.3 Biochemical methods

2.2.3.1 Recombinant protein expression in *E. coli*

Overnight (ON) culture

Heterologous protein expression started by picking a single colony growing on the LB-agar with appropriate antibiotic, and inoculating 5 mL of the LB media with appropriate antibiotic. The culture was incubated overnight at 37˚C at 200 rpm.

Culture conservation

For long-term storage of *E. coli* cells, 500 µL of *E. coli* cells (ON culture) were mixed with 500 µL of a solution of 50 % v/v glycerol (sterilized, reaching the 25 % v/v end concentration) and stored at -80 °C. In case of conservation of the mutant library culture in 96-well microtiter-plates, 100 µL of *E. coli* cells (ON culture) were mixed with 100 µL of a solution of 50 % v/v glycerol (sterilized, reaching the 25 % v/v end concentration) and stored at -80 °C.

Expression of β-etherases and glutathione lyases

Heterologous expression of the β-etherases and glutathione lyases was performed in *E. coli* BL21 (DE3), *E. coli* BL21 Gold (DE3) or *E. coli* C43 (DE3) strains using the T7 expression system providing a high-level expression. T7 RNA polymerase, encoded in the genomic DNA of the *E. coli* strains combined with strong T7 promoter (pET-28a(+), pETDuet-1) or *trc* promoter (pIT2) allows controlled heterologous expression induced by the isopropyl-β-D-thiogalactopyranoside (IPTG).

Heterologous protein expression was generally done in 500 mL TB media (2 L flask) with the respective selection markers. Prepared ON culture was used to inoculate 500 mL of TB media (1 % v/v inoculum). The cells were grown at 37 °C at 200 rpm until they reached OD_{600} of 0.4-0.6 (approx. 2 hours). After the induction with 0.1 mM of IPTG, expression was performed overnight (18 h) at 20˚C (the co-expression experiments were performed at variety of temperatures and lengths (chapter 3.2.2)). After the protein

expression, the cells were harvested by centrifugation at 4,000 rpm for 30 min at 4 °C, cell pellet was washed and stored at -20 °C until further experiments.

2.2.3.2 Enzyme purification

All of the expressed proteins, with intention of use as a purified enzyme, were designed to contain N-terminal hexahistidine tag, allowing relatively simple protein purification by immobilized metal affinity chromatography (IMAC). The principle of the IMAC purification is based on the specific interaction of the histidine to the immobilized metal ions (Ni^{2+}, in our case). Target protein with the polyhistidine tag binds to the Ni^{2+} ions immobilized over NTA (nitrilotriacetic acid) groups on sepharose beads. After the successful binding of the target protein, non specific proteins of the cell were rinsed and the bound desired protein was eluted with the increased imidazole concentration.

Purification using an ÄKTA FPLC system

Cells pellets, prepared as described previously, were resuspended in the binding buffer containing 1 tablet of cOmplete™ Protease Inhibitor Cocktail per 10 mL of the cell mixture. Afterwards, cell mixture was lysed by sonication (5 min of active pulsing, 5 s pulse, 10 s break; at an amplitude of 60 %, on ice). Produced cell lysate was centrifuged (18,000 *g*, 20 min, 4˚C) resulting in separation of cell debris and cell-free extract (CFE). Further, CFE was filtered through a 0.45 µm cellulose acetate membrane filter prior loading it on an equilibrated HisTrap™ FF/HF column. Volume of the column was 5 mL, and the equilibration of the column was performed with 5 column volumes (CV) of binding buffer at a flow rate 1 mL/min. The filtered CFE was loaded to the column with a flow rate of 1 mL/min. The non-specific proteins were eluted by washing the column with 15 CV of equilibration buffer with a flow rate of 1 mL/min. Afterwards, the concentration of the imidazole was gradually increased from 20 to 500 mM by mixing the binding buffer with the elution buffer over 60 minutes with a flow rate of 1 mL/min. Fractions of 1.8 mL were continuously collected and the elution of the target protein was followed by the changes of the intensity of the UV signal at 280 nm. 12 µL samples of the eluted fractions were analyzed by SDS-PAGE (2.2.3.3, SDS-PAGE) to determine the purity of each fraction. Fractions without impurities were pooled together and concentrated via ultrafiltration (Amicon Ultra-15 Centrifugal Filter with 10 kDa cut-off, Merck Millipore, Darmstadt, Germany). PD-10-columns (GE Healthcare) were used to change the buffer from elution to the protein storage buffer. The protein concentration was calculated with the extinction coefficient of the particular purified protein and the absorption at 280 nm measured with the NanoPhotometer NP80. 100 µL aliquots of the protein were stored at -20 °C until further use.

Purification using gravity flow columns

For the purification of the smaller amounts of enzyme or in cases when several proteins had to be purified simultaneously, gravity flow columns were used. 100 mL of the cell culture with expressed protein was reprocessed (harvested, lysed and centrifuged) as before. Filtered CFE was loaded on an equilibrated column containing 2 mL of Sepharose™ 6 Fast Flow resin. After the CFE was loaded on the column, the column was washed with 10 CV of the binding buffer allowing non-specific proteins to wash out. Afterwards, the column was washed with 10 CV of the pre-elution buffer, with increased imidazole concentration, allowing the proteins with naturally high histidine contend to be washed out. And finally, the desired protein was washed out with 10 CV of the elution buffer. Received protein solution was subsequently concentrated, the elution buffer was replaced by the protein storage buffer and the protein was stored as described before.

2.2.3.3 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

Visualization and separation of the protein samples based on the size was performed using SDS-PAGE. Because the structural folding of the protein might influence the migration through the gel, the proteins were denatured with SDS and mercaptoethanol prior to the separation. In the presence of SDS and a mercaptoethanol (cleaving disulfide bonds critical for proper folding) proteins were unfolded into linear chains with negative charge proportional to the molecular weight. Separation of the proteins was performed within the gel of acrylamide and bis-acrylamide, crosslinked by tetramethylethylendiamine (TEMED) and ammonium persulfate (APS). The gel hinders the migration of protein molecules in the electric field depending on their size. As result, the smaller molecules migrate faster than bigger molecules.

For the typical separation, 12 % v/v acrylamide gels were used (composition in **Tab. 2.19**). Prepared samples (either purified protein, CFE, or cell suspension) were mixed with 4x SDS PAGE-sample buffer, heated to 95˚C for 10 min, and centrifuged. Afterwards, 12 µL samples were loaded to the gel and run at 80 V for 10 min to allow the protein sample to penetrate through the stacking gel, followed by approximately 60 min at 120 V. The separation was performed using SDS PAGE-running buffer as electrolyte. After the protein separation, the gel was stained in the SDS PAGE-gel staining solution for 3 hours. Subsequently the gel was destained using SDS PAGE-gel destaining solution, preferably over night.

Tab. 2.19: Composition of the two gels that creates a 12 % v/v acrylamide SDS-PAGE gel.

Component	Stacking gel	Running gel
MilliQ water	1500 µL	2250 µL
Acrylamide (40 %)	375 µL	1500 µL
1.5 M Tris-HCL (pH 8.8), 0.4 % SDS	-	1250 µL
0.5 M Tris-HCL (pH 6.8), 0.4 % SDS	625 µL	-
10 % w/v APS	25 µL	50 µL
TEMED	2.5 µL	5 µL

2.2.4 Enzyme characterization

2.2.4.1 Activity assays

Specific activity of an enzyme is an important characteristic used to describe and compare enzymes catalyzing the same reaction. The measure of catalytic ability represented by enzyme activity can be quantified by the decrease in substrate concentration in a period of time or by the increase in concentration of a product after a period of time. For the characterization of β-etherases and glutathione lyases, the 1 mL reactions were performed.

β-Etherase activity assay

The β-etherase activity of each purified enzyme towards **1** and **2** was determined by quantifying the amount of released 2,6-dimethoxyphenol by HPLC. The quantification of the 2,6-DMP was performed by using calibration curve of respective commercial standard (Sigma–Aldrich). The 1 mL assay mixture contained 100 mM glycine/NaOH buffer, pH 9.5; 0.2 mM of substrate (dissolved in DMSO, final DMSO concentration, 5 % v/v); 1 mM of reduced GSH (5 eq. to the substrate), and 0.5-10 µg of purified β-etherase. Reactions were carried out at 25˚C and stopped by addition of 0.33 volume equivalents of 3.55 M H_2SO_4 (final H_2SO_4 concentration, 0.89 M) after different incubation times. Precipitated protein was removed by centrifugation (15,000 *g* for 5 min), and the supernatant was analyzed on HPLC system. HPLC chromatogram is shown in chapter 7.7. One unit was defined as the amount of enzyme that converts 1 µmol of substrate per min. Specific activity was expressed as milliunits per milligram of protein.

Ability of β-etherases to cleave ether bond in non-lignin substrates **30**, **31**, **32**, **33**, and **34** was determined by decrease of peaks and formation of novel peaks observed on HPLC chromatograms. The 1 mL assay contained (10 reactions were prepared for each enzyme + 5 control, non-enzymatic reactions) 100 mM phosphate buffer, pH 8; 0.2 mM of substrate (dissolved in DMSO, final DMSO concentration, 10 % v/v); 1 mM of reduced GSH (5 eq. to the substrate), and 50 µg of purified β-etherase. Reactions were carried out at 25˚C with shaking 800 rpm. In the desired time point, whole volume of two

enzymatic reactions and one control reaction were stopped by dilution with DMSO in 1:10 ratio to assure complete dissolving of the substrate and possible product/s. Precipitated protein was removed by centrifugation (15,000 g for 5 min), and the supernatant was analyzed on HPLC system.

Glutathione lyase activity assay

The glutathione lyase activity of each purified enzyme was determined by quantifying the amount of released veratryl glycerol (VG) or by quantifying the amount of released 3,4-dimethoxyacetophenone (3,4-DAP) by HPLC. The quantification of VG was performed by using the calibration curve of enzymatically prepared VG (extracted and purified) and quantification of 3,4-DAP was performed by using the calibration curve of respective commercial standard (Sigma–Aldrich). The 1 mL assay mixture contained 100 mM glycine/NaOH buffer, pH 9.0; 0.2-0.4 mM of substrate; 1-2 mM of reduced GSH (5 eq. to the substrate), and 0.1-5 µg of purified glutathione lyase. Reactions were carried out at 25˚C and stopped by addition of 0.33 volume equivalents of 3.55 M H_2SO_4 (final H_2SO_4 concentration, 0.89 M) after different incubation times. Precipitated protein was removed by centrifugation (15,000 *g* for 5 min), and the supernatant was analyzed the HPLC system. HPLC chromatogram is shown in chapter 7.7. One unit was defined as the amount of enzyme that converts 1 µmol of substrate per min. Specific activity was expressed as units per milligram of protein.

Vanillin release assay and enzyme kinetics

The β-etherase activity of each purified enzyme (or prepared CFE) towards **3** was determined by continuous following of released vanillin by spectrophotometer. The quantification of the vanillin was performed using calibration curve of respective commercial standard (Sigma–Aldrich). The reactions were performed in 96-well microtiter-plates. The 200 µL assay mixture contained 100 mM glycine/NaOH buffer, pH 9.5; diverse substrate **3** concentration (0.05, 0.075, 0.1, 0.2, 0.3, 0.4, 0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0, 2.5, 3.0, 3.5 and 4.0 mM) while the DMSO, in which the substrate was dissolved, was kept at 10 % v/v. The concentration of the GSH was kept at 15 mM to ensure the saturation and the amount of each enzyme was kept constant (1-50 µg), while in case of CFE, 10 µL was used for the reaction. The reaction progress was monitored by absorption at 360 nm every 8 s for the 81 s reaction. The slopes from the measured reaction points were used in calculations. Relation between VG concentration and spectrophotometric response was calculated based on the calibration curve shown in chapter 7.11. One unit was defined as the amount of enzyme that converts 1 µmol of substrate per min. Specific activity was expressed as milliunits per milligram of protein.

The kinetics parameters for β-etherases were determined by fitting the experimental data

with either the Michaelis-Menten equation (**Eq. 1**) or the Hill's equation (**Eq. 2**) to include cooperative bindings in the fitting. Reaction rates (V) at different substrate concentrations ([S]) were measured in order to determine K_M (substrate concentration at which the reaction rate is at half-maximum), V_{max} (maximum rate at a given enzyme concentration), and k_{cat} (the number of substrate molecule each enzyme site converts to product per unit time).

$$V = \frac{V_{max} \cdot [S]}{K_m + [S]} \qquad (\boldsymbol{Eq.1}) \qquad V = \frac{V_{max} \cdot [S]^n}{(K_{0.5})^n + [S]^n} \qquad (\boldsymbol{Eq.2})$$

The kinetic parameters relative to β-etherases and substrate **3** were calculated by keeping the enzyme and GSH concentrations constant and varying the substrate concentration. Kinetic analyses of all prepared mutants in this thesis are shown in chapter 7.8.

2.2.4.2 Selectivity assays

E-value

The selectivity of an enzymes is commonly quantified by the E-value. The E-value describes how effectively/preferentially the enzyme converts one enantiomer of a racemic mixture of a substrate. An E-value of a non-selective enzyme is described by value 1. The E-value of a strictly selective enzyme (only one enantiomer is converted) displays the E-value of ≥200. Due to the exponential increase of the analytical error in logarithmic formula, the E-values higher than 200 are not applied [130]. The selectivity of the β-etherases was determined by using chiral HPLC analysis with a Chiralcel OD-RH column (4.6 × 150 mm). Biocatalytic reactions were performed as described in 2.2.4.1 (β-etherases activity assay). The selectivity was calculated based on the decrease of the each 2,6-MP-VG enantiomers according to **Eq. 3** [130].

$$\mathrm{E} = \frac{ln[(1-c)(1-e.e_S)]}{ln[(1-c)(1+e.e_S)]} \qquad (\boldsymbol{Eq.3})$$

R/S ratio

The selectivity of the glutathione lyases could not be calculated via E-value, since the enantiomers of the racemic substrate could not be separated by HPLC. Therefore, the specific activity towards each enantiomer was determined as described in 2.2.4.1 (glutathione lyases activity assay) and compared.

2.2.4.3 Tryptophan fluorescence assay

Changes of the intrinsic tryptophan fluorescence upon titration with substrates **1**, **2**, **4**, **5**, and **6** were followed on a CLARIOstar microtiter plate reader. The reactions were

performed in 96-well microtiter-plates for fluorescence measurement (black). The starting volume of 200 µL consisted of the enzyme solution with concentration of 500 µg/mL in 100 mM glycine buffer, pH 9.0. Emission spectra from 320 to 360 nm were recorded after excitation at 295 nm. The maximum emission wavelength of the intrinsic fluorescence of the protein was determined to be at 340.6 nm. Based on this, the change in intrinsic fluorescence intensity during titration with the ligand stock (10 mM, solved in DMSO) was measured at 340.6 nm. After each ligand titration, the fluorescence was measured. Values were corrected for dilution and also the changes of the fluorescence intensity caused by the DMSO (without ligand) were subtracted. Apparent binding constants were obtained by nonlinear regression. The fluorescence intensities were measured over a range of ligand concentrations and were subsequently fitted by nonlinear regression using **Eq. 4**, in which F is the observed relative fluorescence intensity, F_0 the fluorescence intensity of the enzyme without ligand present, ΔF_{max} fluorescence change, [X] the ligand concentration, and K_D the binding constant.

$$F = \left[1 - \left(F_0 + \frac{\Delta F_{max} \cdot [X]}{K_D + [X]}\right)\right] \cdot 100\ \% \qquad (\boldsymbol{Eq.4})$$

2.2.4.4 Isothermal titration calorimetry (ITC)

Interactions between protein and ligand to determine the thermodynamic parameters in solution was performed by ITC using an MCS-ITC calorimeter. The cofactor reduced glutathione (GSH) was dialyzed against the LigF protein or against the LigF protein with the various substrate **4** concentration bound within the protein. Both, the GSH and proteins solution were solved with 100 mM HEPES with 10 % v/v methanol, pH 8. During the titration, 5 mM GSH was continuously injected into 0.1 mM of LigF protein at 25˚C in 14 µL steps up to an 8-fold molar excess. The mixing heat of the GSH, measured by injecting GSH into buffer, was subtracted from the signals obtained from the binding reactions. ITC data of two independent measurements were analyzed using Origin V7.0 with MicroCal ITC add-on.

2.2.5 Whole-cell reactions

2.2.5.1 Analytical-scale reactions

1 mL reactions containing 100 mM potassium phosphate buffer, pH 8, 0.2 mM substrate **1** (dissolved in an organic solvent-water mixture (1:1) with a final organic solvent concentration of 0.5 % v/v in the reaction) and *E. coli* cells at an OD_{600} of 40 were performed at 25°C with shaking at 800 rpm. Samples were taken after different time intervals. Reactions were stopped by adding 0.33 volume equivalents of 3.55 M H_2SO_4 (final H_2SO_4 concentration, 0.89 M), resulting also in partial cell lysis. Cells and cell debris were removed by centrifugation (17000 *g*, 10 min) before HPLC analysis.

Conversions for optimization of enzyme expression were performed with cells at an OD_{600} of 10. Conversions for pH optimization were carried out in different reaction buffers with pH values of 5 - 11 using also an OD_{600} of 10. To study the impact of substrate concentration on the reaction rate, **1** was added in varying concentrations between 0.2 and 10.0 mM, while the organic solvent concentration was fixed at 5 % v/v. Impact of the co-solvent and its concentration was tested using DMSO, isopropyl alcohol and methanol. Solvent concentrations (final concentration in the reaction mixture) were tested at 0.5 %, 2.5 %, 5.0 %, 7.5 %, 10.0 %, 12.5 % v/v, while the substrate concentration was fixed at 0.2 mM. Additionally, reactions without co-solvent were performed at substrate concentrations of 0.2 mM and 1.0 mM. To check the influence of glucose addition on the conversion of 0.8 and 10 mM 2,6-MP-VG (dissolved in methanol, 5 % v/v final concentration in the reaction mixture), each 10 mM glucose was added to the reaction buffer. Increased substrate and solvent concentrations were performed in conversion of 10 mM 2,6-MP-VG with DMSO concentration of 5 %, 15 % and 25 % v/v.

Reactions with recycling of the cell were performed at 5 mL volume containing cell at an OD_{600} of 40, with the substrate concentration fixed at 0.8 mM of the 2,6-MP-VG dissolved in DMSO or isopropyl alcohol or methanol, while the solvent concentration was kept at 5 % v/v. In these reactions, the impact of the glucose was also tested by addition of 10 mM of glucose to the reaction. Reaction samples were taken after 4 hours and were reprocessed as described before. Remaining reaction mixture was centrifuged at 6,000 *g* for 5 min. The supernatant was discarded and the cells were resuspended in the fresh medium containing new substrate (correlated to the volume of the sample taken). The cell recycle was repeated for four cycles.

For comparison of kinetic resolutions using either *E. coli* C43 (DE3) cells harboring only LigE, *E. coli* C43 (DE3) cells harboring LigE and LigG-TD, or a combination of *E. coli* C43 (DE3) cells harboring LigE and *E. coli* C43 (DE3) cells containing LigG-TD to obtain

(*S*)-2,6-MP selectively, reactions were performed at an OD_{600} of 40 for each strain. The substrate concentration was set to 10.0 mM dissolved in methanol (final solvent concentration of 5 % v/v). Likewise, kinetic resolutions using either *E. coli* C43 (DE3) cells harboring only LigF-NA, *E. coli* C43 (DE3) cells containing LigF-NA and LigG-TD, or a combination of *E. coli* C43 (DE3) cells harboring LigF-NA and *E. coli* C43 (DE3) cells harboring LigG-TD to obtain (*R*)-2,6-MP selectively, were performed using the same reaction conditions. Relations between OD_{600} and dry cell weight was calculated based on calibration curve shown in chapter 7.11.

2.2.5.2 Semi-preparative-scale reactions

To prepare enantiopure (*S*)-2,6-MP-VG and (*R*)-2,6-MP-VG, semi-preparative-scale conversions were performed in 30 mL of 100 mM potassium phosphate buffer, pH 8, with 10 mM substrate 2,6-MP-VG dissolved in methanol (final solvent concentration of 5 % v/v) and *E. coli* C43 (DE3) cells containing either LigE and LigG-TD at an OD_{600} of 80 or LigF-NA and LigG-TD at an OD_{600} of 40. Reactions were incubated at room temperature with shaking (800 rpm) and stopped by adding 0.33 volume equivalents of 3.55 M H_2SO_4 (final H_2SO_4 concentration, 0.89 M) after 46 hours. Reaction mixtures were diluted in 1:9 ratio using methanol. Cells were removed by centrifugation (3,488 *g*, 40 min) and the resulting supernatant was used for achiral and chiral HPLC analysis. Afterwards, the excess of methanol was removed again by evaporation to reduce the reaction volume and products were extracted from the water fraction with ethyl acetate (2x30 mL). The organic phases were combined and the ethyl acetate was removed by evaporation. The desired enantiomers (*S*)-2,6-MP-VG and (*R*)-2,6-MP-VG were purified by column chromatography (60M FLASH, 0.04-0.063 mm) using a mixture of dichloromethane and ethyl acetate in 2:1 ratio as solvent. 43 mg (80 % yield) of (*S*)-2,6-MP-VG and 39 mg (72 % yield) of (*R*)-2,6-MP-VG were obtained as pure compounds. The enantiomers were dissolved in dichloromethane (*c* = 20 mg/mL) to determine their optical rotation on a Propol Digital Automatic Polarimeter (Dr. Kernchen, Seelze, Germany). Obtained specific rotation values for (*S*)-2,6-MP-VG and (*R*)-2,6-MP-VG are $[\alpha]^{20}_D = -21.85°$ and $[\alpha]^{20}_D = +21.15°$, respectively. NMR data of both enantiomers were consistent with literature [110].

2.2.6 Protein engineering

2.2.6.1 Generation of mutant libraries

The site saturation random mutagenesis libraries were prepared via PCR reaction using degenerated primers (2.2.2.4, QuikChange® PCR). Afterwards, the chemocompetent *E. coli* DH5α cells were transformed (2.2.2.3, Transformation of the chemically competent cells), and the multiple cell colonies growing on LB agar with appropriate antibiotic were combined (mixed with 500 µL LB media) and the mutant library was isolated in form of plasmids (2.2.2.6). With isolated plasmid DNA of the library, the expressing strain *E. coli* BL21 (DE3) Gold was transformed. The single colonies were subsequently picked and transferred into the 96-well microtiter-plate containing 200 µL LB media with appropriate antibiotic. The cells were cultivated overnight at 37˚, 800 rpm. The cryo culture of the mutant library was prepared as described at 2.2.3.1 (Culture conservation) and sealed with SILVERSEAL SEALER Aluminium foil (Greiner Bio-One, Austria). The quality of the library was tested by sequencing of the isolated library plasmid (2.2.2.6).

2.2.6.2 Recombinant protein expression in 96-well microtiter-plate format

The expression of the created mutant libraries was performed in 96-well microtiter plates. The 200 µL of LB media with appropriate antibiotic and 0.1 mM IPTG was inoculated with 5 µL of the prepared ON-culture. The micro-titer plate was incubated for 24 h at 20˚C with 800 rpm. During the time of the protein expression, the micro-titer plate was sealed with a breathable rayon film seal for biological cultures (VWR, USA) to prevent media evaporation and cross-contamination. Afterwards, the final OD_{600} of the cultures was measured on CLARIOstar microtiter plate reader by diluting the cell cultures in ratio 1:20 with the LB media. The cells were harvested by centrifugation at 4000 *g*, 4˚C for 20 min and stored at -20˚C until further use.

2.2.6.3 Mutant library analysis

The harvested cells from the mutant library were resuspended in 200 µL of B-PER (bacterial protein extraction reagent). The cells were lysed at room temperature while shaken at 800 rpm for 1 h. After the lysis, the cell lysate was centrifuged (4000 *g,* 20 min, 4 °C), the supernatant was diluted 1:50 with storing buffer, and the prepared CFE was used for reaction. Analysis of the prepared mutant library was tested using vanillin release assay as described before (2.2.4.1 Vanillin release assay). However, in this case, the length of the reaction was extended to 30 min.

2.2.6.4 Validation of interesting mutants

Single mutants with interesting properties (increased activity, loss of activity) exposed by vanillin assay were validated by expressing the enzyme in 100 mL of TB media as described at 2.2.3.1 (β-etherases and glutathione lyases expression) and purified using gravity flow columns (2.2.3.2, Purification using gravity flow column). Prepared purified enzyme was subsequently further analyzed by HPLC.

2.2.7 Analytical methods

2.2.7.1 High-performance liquid chromatography (HPLC)

HPLC is an important analytical method regularly used for separation and quantification of liquid samples. In this method, the sample is pumped by the mobile phase through a column that contains a stationary phase. Different interactions between sample components and the stationary phase, cause the components to move through the column with different average velocities and thus separating the components. One of the benefits of the HPLC method is that it is versatile and variety of modes are possible. The mode in which the mobile phase is polar and stationary phase is non-polar is called reverse-phase HPLC. In this thesis, the reverse-phase chromatography, where stationary phase consisted of small silica particles whereas the mobile phase consisted of mixture of polar solvents, was used.

The samples containing substrates **1**, **2**, **4**, **5**, **6**, **7**, and **8** were analyzed with Nexera XR HPLC System with a Nucleosil® 100-5 C18 column (4.6 x 250 mm) as a stationary phase. The mobile phase consisted of an isocratic mixture of water, acetonitrile and trifluoroacetic acid (TFA) in volume ratio 49.95:50:0.05 with a flow rate of 1 mL/min.

The samples containing non-lignin models **30**, **31**, **32**, **33**, and **34** were analyzed with Nexera XR HPLC System with a Nucleosil® 100-5 C18 column (4.6 x 250 mm) as a stationary phase operated with the flow rate of 1 mL/min using binary gradient of A (water and 0.1 % TFA) and B (acetonitrile) described on the **Fig. 2.2**.

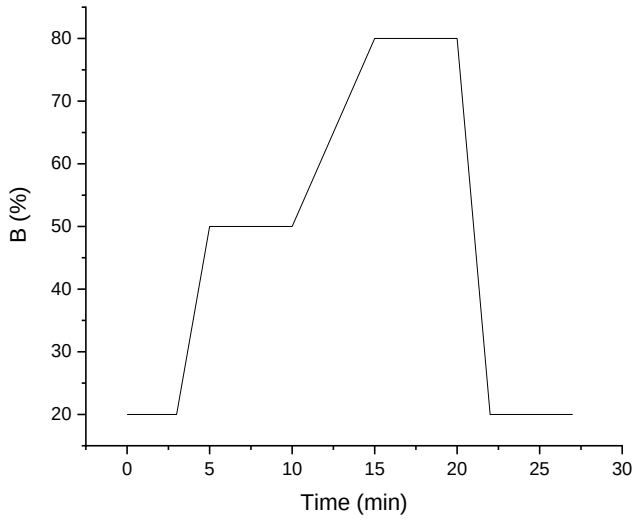


Fig. 2.2: Binary gradient profile of A (water and 0.1 % TFA) and B (acetonitrile) of the analytical method for the analysis of substrates **30**, **31**, **32**, **33**, and **34**.

The chiral analysis was performed using Chiralcel OD-RH column (4.6 × 150 mm) as a stationary phase. The mobile phase consisted of isocratic mixture of water and acetonitrile in volume ratio 70:30 with a flow rate of 1 mL/min.

2.2.7.2 Liquid chromatography–mass spectrometry

LC-MS analyses were performed by Ulrike Beutling in cooperation with the Department of the Chemical Biology at the HZI Braunschweig in the group of Prof. Dr. Mark Brönstrup. For liquid chromatography, UHPLC-system a Ultimate3000RS from Dionex/Thermo (with autosampler, binary high gradient pump, column oven, 6-port-column-switching-option, DAD-detector) was used. The samples were separated by column Kinetex 1,7µ C18 100A, 150x2,1 mm from Phenomenex as stationary phase operated with the flow rate of 300 µL/min using binary gradient of solvent A (water with 0,1 % formic acid) and solvent B (acetonitrile with 0,1 % formic acid) at 40°C. Method gradient: 0-2 min - 1 % of solvent B, 2-20 min - 100 % of solvent B. After this analysis, the column was washed with 100 % B and returned to the starting conditions with 1 % of solvent B. Overall run time was 30 min.

For the mass spectrometry, mass spectrometer Bruker maxis HD UHR-TOF equipped with source Apollo II Elektrospray was used. Within the first 0,3 min of the run a Na-Formiat-Cluster as calibrant was infused to the system (for internal calibration). For the lock mass calibration with 622 m/z for positive and 556 m/z for negative ion mode Hexakis (2,2-difluoroethoxy) phosphazene was used. Calibrations were done in Data

Analysis software. Parameters for MS acquisition were as follows: source type: ESI; scan range: 50 – 1500 m/z; ion polarity: positive or negative; capillary voltage: 4500 V; nebulizer pressure: 4.0 bar; dry heater: 200°C, and dry gas: 9.0 L/min.

2.2.7.3 Nuclear magnetic resonance

NMR analyses were performed in the central facilities of the TU Braunschweig using Bruker Avance AVII 600 NMR spectrometer.

2.2.7.4 Mass spectrometry

MS analyses were performed in the central facilities of the TU Braunschweig by Dr. Uli Papke.

2.2.8 Bioinformatic methods

2.2.8.1 BLAST

BLAST (Basic Local Alignment Search Tool) is an alignment based online tool provided by NCBI to align, compare and search for regions of local similarity between sequences. The program is able to compare nucleotide or protein sequences to big sequence databases and calculate the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families. BLAST tool was used to search for the sequence-wise related LigE and LigF homologs. Gathered sequences were align in multiple sequence alignment (MSA) using the free online tool BioEdit and are shown in chapters 7.9 and 7.10.

3 Results

3.1 Production and analysis of recombinant β-etherases and glutathione lyases

3.1.1 Heterologous production and purification of β-etherases and glutathione lyases

The presented thesis is the continuation of the research performed by the Pere Picart that is summarized in excellent publications [105,110]. Picart *et al*. [105,110] performed the cloning of the β-etherases and glutathione lyases into pET28a(+) vectors. β-Etherases that are used in this project are: LigE, LigF, LigP from *Sphingobium paucimobilis* SYK-6; LigE-NS, LigF-NS from *Novosphingobium* sp. PP1Y; and LigE-NA, LigF-NA from *Novosphingobium aromaticivorans* DSM1244. Glutathione lyases used in this thesis are: LigG from *Sphingobium paucimobilis* SYK-6; LigG-NS form *Novosphingobium* sp. PP1Y; and LigG-TD from *Thiobacillus denitrificans* ATCC 25259.

All of the above-mentioned enzymes were produced in *E. coli* BL21 (DE3) Gold. Cells were grown in TB media containing 50 µg/mL of kanamycin at 37˚C, 220 rpm until the optical density of the culture at 600 nm (OD_{600}) reached 0.4-0.6. At that point, expression was induced using 0.1 mM of IPTG and the temperature was reduced to 20˚C for 18-22 hours followed by harvesting by centrifugation. Since all of the enzymes carried an N-terminal His-tag, the purification was carried out via immobilized metal affinity chromatography (IMAC) using a Ni-NTA column connected to an ÄKTA FPLC system. As an example, chromatograms of the purification process and the SDS-PAGE of the collected fractions of LigE are shown in **Fig. 3.1**. In order to elute the his-tagged protein from the Ni-NTA column, the imidazole concentration in the buffer was gradually increased. Proteins with no His-tag were eluted directly after the column loading, while proteins with naturally high histidine concentration were eluted at the beginning of the imidazole gradient.

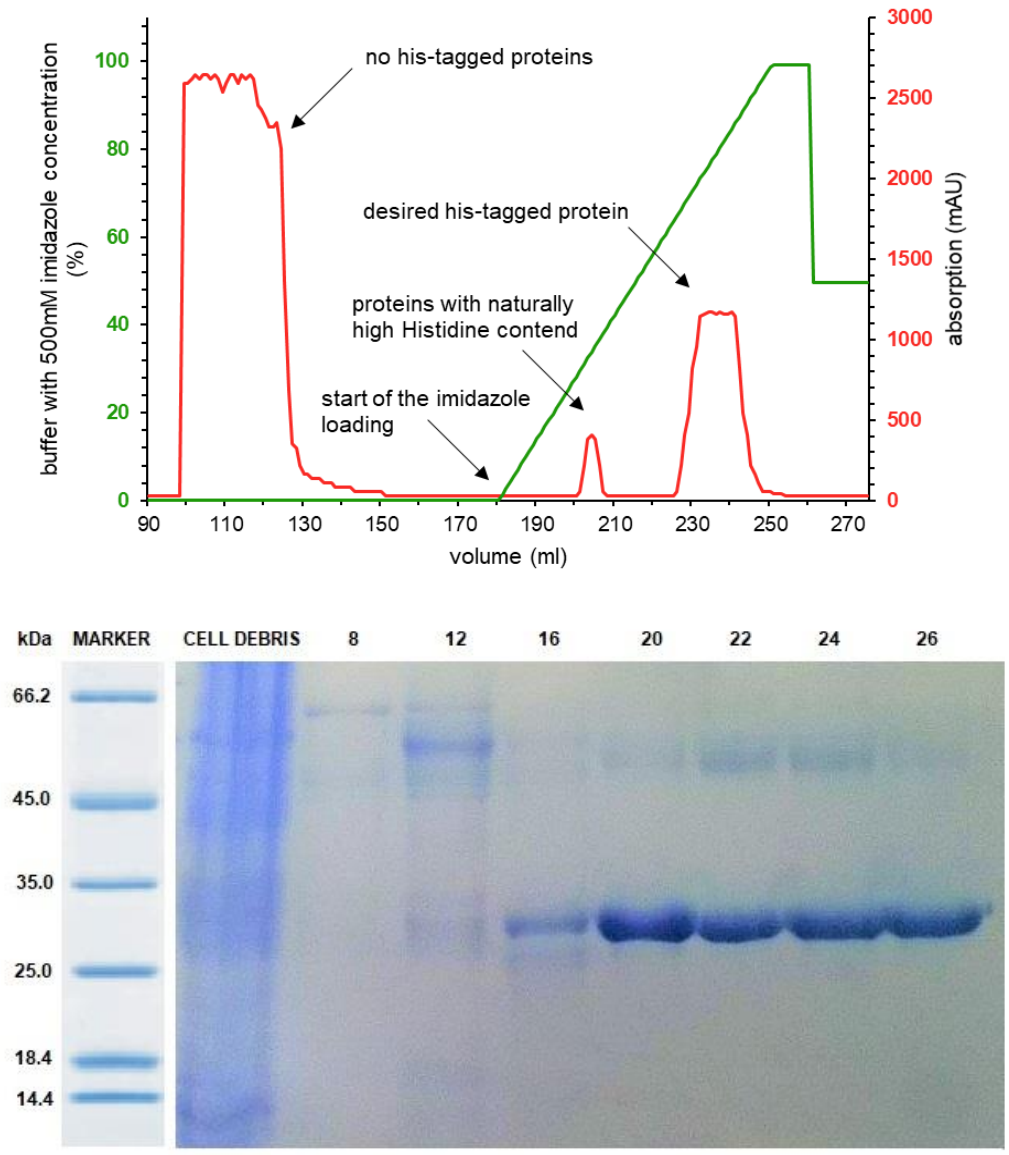


Fig. 3.1: Top: Chromatogram for the purification of His-tagged LigE with the red line indicating protein elution via absorbance at 280 nm and the green line representing the imidazole concentration in the buffer. Bottom: SDS-PAGE analysis of fractions from LigE purification [lane 1: protein marker, lane 2: diluted cell debris after sonification and centrifugation, lanes 3-9: different fractions from the imidazole gradient for protein elution]. Fractions 20-26 show target protein LigE with a size of ~30kDa. Marker: PageRuler Prestained Ladder (Life Technologies).

The fractions showing a clear band at the size of the desired protein were collected, concentrated and desalted. The yields of each of the purified proteins were calculated, and are summarized in **Tab. 3.1**.

Tab. 3.1: Protein yields (in $mg_{PROTEIN}/L_{MEDIA}$) of β-etherases and glutathione lyases after affinity purification and their molecular weight.

Enzyme	Molecular weight (kDa) including His-tag	Yield (mg/L)
LigE	34.2	40
LigE-NS	33.0	130
LigE-NA	33.3	76
LigP	31.0	52
LigF	31.8	67
LigF-NS	30.8	146
LigF-NA	30.6	52
LigG	32.4	38
LigG-NS	32.0	85
LigG-TD	32.9	73

Yields of the proteins ranged between 38 $mg_{PROTEIN}/L_{MEDIA}$ and 146 $mg_{PROTEIN}/L_{MEDIA}$. When adding 2 mM $MgSO_4$ to the cultivation media, the yield of LigE and LigF could be increased by 56 % and 190 %, respectively [270].

During the purification process of LigF-NA, increased proteolytic activity was observed despite the fact that *E. coli* BL21 (DE3) Gold lacks the Lon and OmpT proteases which can degrade recombinantly produced proteins. Proteolytic digestion of LigF-NA was observed by formation of two distinct protein bands of similar size on the respective SDS-PAGE gel (**Fig. 3.2**). Due to this, 10 µM phenylmethylsulfonyl fluoride (PSMF) and cOmplete™ Protease Inhibitor Cocktail were tested as protease inhibitors during cell disruption and purification. While using PMSF did not show any improvement, cOmplete™ Protease Inhibitor Cocktail prevented LigF-NA degradation. Therefore, the use of the cOmplete™ Protease Inhibitor Cocktail was standardly implemented into the purification protocol.

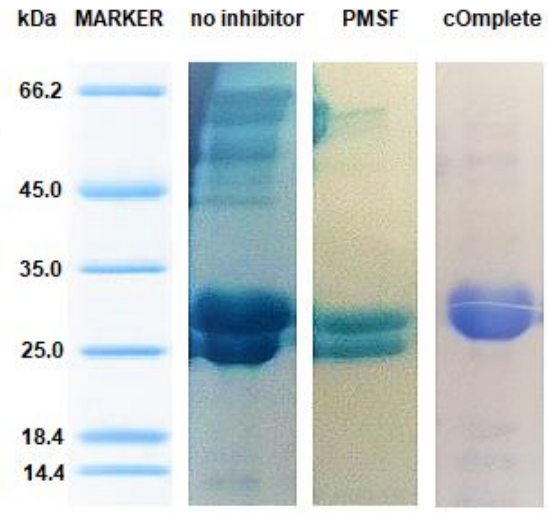


Fig. 3.2: SDS-PAGE gel showing fractions of purified LigF-NA [lane 1: protein marker, lane 2: purified LigF-NA without addition of protease inhibitor, lane 3: purified LigF-NA with 10 µM of PMSF, lane 4: purified LigF-NA with cOmplete™]. In lanes 2 and 3 two bands at ~25kDa and ~27kDa are visible, suggesting degradation of the recombinant protein. Lane 4 shows target protein LigF-NA with a size of ~30kDa as one band. Marker: PageRuler Prestained Ladder (Life Technologies).

3.1.2 Characterization of β-etherases

β-Etherases were previously shown to cleave the β-*O*-4 aryl ether bond in different lignin model substrates as well as non-lignin ether compounds, which revealed a significant influence of the side chain as well as different substitution patterns on the aromatic rings on enzyme activity [104,107–110,112,131]. Model substrates used in this work for the characterization of β-etherases at optimal pH and temperature are shown in **Fig. 3.3**.

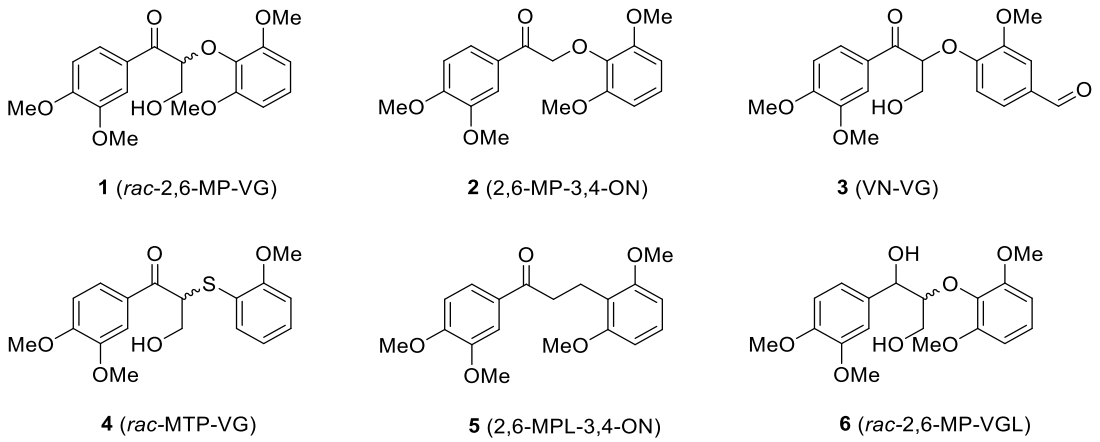


Fig. 3.3: Lignin model substrates and non-lignin ether compounds used in this work for characterization of β-etherases. [(**1**, *rac*-2,6-MP-VG): β-(2,6-dimethoxyphenoxy)-α-veratrylglycerone; (**2**, 2,6-MP-3,4-ON): β-(2,6-dimethoxyphenoxy)-α-(3,4-dimethoxyphenyl)-ethan-1-on; (**3**, VN-VG): β-vanillinyl-α-veratrylglycerone; (**4**, *rac*-MTP-VG): β-(2-methoxytiophenoxy)-α-veratrylglycerone; (**5**, 2,6-MPL-3,4-ON): β-(2,6-dimethoxyphenyl)-α-(3,4-dimethoxyphenyl)-ethan-1-on; (**6**, *rac*-2,6-MP-VGL): β-(2,6-dimethoxyphenoxy)-α-veratrylglycerol].

Substrate **1** represents the standard lignin model compound of β-etherases, for which activities have been determined previously [110]. On the other hand, the corresponding achiral, side-chain-truncated model substrate **2** had not been tested before. Substrates **4**, **5**, and **6** represent variations of model compound **1** where the aryl ether bond was replaced by a thioether bond or a carbon-carbon bond and the Cα carbonyl group was replaced by a hydroxyl group, respectively. Reactions with substrates **1**, **2**, **4** and **6** were analyzed by HPLC. In contrast, substrate **3** represents a model substrate suitable for spectrophotometric analysis. Hence, activity of the β-etherases towards substrate **3** was determined spectrophotometrically by measuring the released product, vanillin, at 360 nm. Enantioselectivity of the enzymes in the kinetic resolution of **1** was analyzed by chiral HPLC which allowed separation of substrate enantiomers (*R*)-**1** and (*S*)-**1**. Specific activities of all β-etherases in the conversion of substrates **1**-**6** as well as their enantioselectivities are summarized in **Tab. 3.2**.

Tab. 3.2: Specific activities of β-etherases in the conversion of lignin model substrates **1**-**6** and their enantioselectivities in the kinetic resolution of **1**.

Enzyme	Specific activity towards (U/mg)						Stereoselectivity	E-value*
	1	**2**	**3**	**4**	**5**	**6**		
LigE	1.75	0.34	6.88	n.d.	n.d.	n.d.	(*R*)-selective	>200
LigE-NS	1.00	0.30	3.44	n.d.	n.d.	n.d.	(*R*)-selective	>200
LigE-NA	0.73	0.05	2.32	n.d.	n.d.	n.d.	(*R*)-selective	>200
LigP	0.05	n.d.	n.d.	n.d.	n.d.	n.d.	(*R*)-selective	>200
LigF	1.26	0.70	2.19	n.d.	n.d.	n.d.	(*S*)-selective	>200
LigF-NS	1.60	1.30	1.43	n.d.	n.d.	n.d.	(*S*)-selective	>200
LigF-NA	2.40	3.20	3.17	n.d.	n.d.	n.d.	(*S*)-selective	>200

n.d.= not detected, 1U= 1 µmol/min, * calculated according to Chen *et al*. [130].

In agreement with previously published information [107], the tested β-etherases display absolute enantioselectivity towards chiral substrate **1** with all E-type etherases exhibiting (*R*)-selectivity while F-type enzymes show (*S*)-selectivity. From the (*R*)-selective β-etherases, LigE displays the highest specific activity, while LigF-NA is the most active among the (*S*)-selective β-etherases. Absolute activity values of β-etherases, however, differ from the ones published by Picart *et al*. [110]. This is likely explained by the use of differing enzyme and substrate concentrations in reactions as well as different time points of sample drawing from the reactions, which will have an impact on the final activity values. A very similar trend for the specific activities of the β-etherases was observed with substrates **2** and **3,** where LigE and LigF-NA were the most active (*R*)- and (*S*)-selective enzymes, respectively. Substrates **2** and **3**, however, were not converted by LigP. Neither of the produced β-etherases were able to convert substrates **4** and **5**, in agreement with their specificity for ether bond cleavage. DFT calculations carried out by the group of Prof. Dr. Christoph Jacob (Theoretical Chemistry, TU Braunschweig)

suggested that the cleavage of the thioether bond in **4** would be an endothermic reaction (compared to an exothermic reaction for conversion of the corresponding ether compound), which explains the observed lack of enzyme activity. Moreover, the substrate specificity of β-etherases was reported to be limited to Cα-carbonyl-containing lignin model substrates, whereas the corresponding hydroxyl-containing substrates are not converted [98–101,110]. This could be confirmed in this work as no activity towards substrate **6** could be detected.

To further analyze and compare the catalytic performance of the different β-etherases, kinetic studies using **3** as a substrate were performed to determine kinetic parameters K_M and k_{cat} spectrophotometrically (**Tab. 3.3**).

Tab. 3.3: Kinetic parameters, K_M and k_{cat}, of studied β-etherases in the conversion of **3** determined at optimal pH, 20˚C, and with an excess of GSH. For LigF, substrate inhibition was observed.

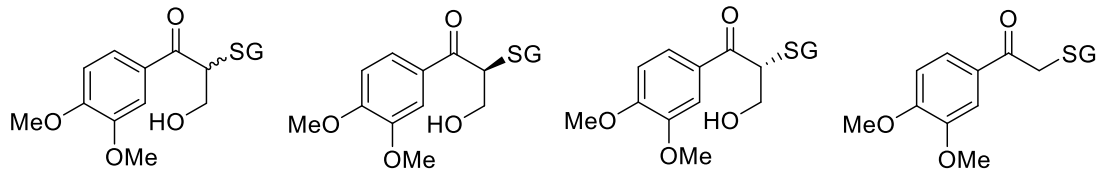
Enzyme	K_M (mM)	K_I (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (mM^{-1} s^{-1})
LigE	0.1±0.02	-	6.2±0.15	62
LigE-NS	0.3±0.03	-	5.3±0.18	18
LigE-NA	0.8±0.20	-	5.0±0.39	6.3
LigF	0.9±0.16	3.6±0.97	4.8±0.56	5.3
LigF-NS	0.7±0.07	-	3.5±0.11	5.0
LigF-NA	0.4±0.07	-	6.0±0.26	15

Similar to the specific activities of (*R*)-selective β-etherases, LigE exhibits the highest k_{cat}. The turnover number of the LigE-NS and LigE-NA, compared to of LigE, display 15 % and 20 % lower value, respectively. Moreover, the K_M value of LigE is 3.5 and 7.5 times lower than those for LigE-NS and LigE-NA, resulting in the highest catalytic efficiency of 62 mM^{-1} s^{-1} for LigE among the tested enzymes. Among the (*S*)-selective β-etherases, LigF was found to display substrate inhibition with **3**. This significantly affected the K_M value of LigF for **3**, which was ca. 1.5-fold and ca. 2-fold higher than that of LigF-NS and LigF-NA, respectively. The highest turnover number of (*S*)-selective β-etherase was exhibited by LigF-NA. In contrast, the k_{cat} of LigF and LigF-NS is ca. 20 % and ca. 40 % lower, respectively, when compared to LigF-NA. The most active (*S*)-selective β-etherase (LigF-NA) showed the K_M value of 434 µM, which is ca. 2.0 and 1.5 times lower than those of LigF and LigF-NS, ensuring the highest catalytic efficiency of 15 mM^{-1} s^{-1} for LigF-NA. Overall, this kinetic study confirmed that enzymes LigE and LigF-NA are the most efficient (*R*)-selective and (*S*)-selective β-etherases for conversion of the tested lignin model substrates.

3.1.3 Characterization of glutathione lyases

β-Etherases catalyze the nucleophilic attack of GSH on the β-position of model substrates **1** and **2** releasing 2,6-dimethoxyphenol (2,6-DMP) and a GSH-conjugated

aromatic compound as products. Subsequently, glutathione lyases catalyze the further conversion of the GSH adduct with the help of another GSH molecule resulting in the release of GSSG and the final product. In this matter, various GSH-conjugates were tested in order to characterize the produced glutathione lyases (**Fig. 3.4**).



rac-**11** (GS-VG) *(R)*-**11** (*(R)*-GS-VG) *(S)*-**11** (*(S)*-GS-VG) **12** (GS-3,5-ON)

Fig. 3.4: Model substrates used in this work for characterization of available glutathione lyases. [(*rac*-**11**, GS-VG): β-glutathionyl-α-veratrylglycerone; ((*R*)-**11**, (*R*)-GS-VG): (*R*)-β-glutathionyl-α-veratrylglycerone; ((*S*)-**11**, (*S*)-GS-VG): (*S*)-β-glutathionyl-α-veratrylglycerone; (**12**, GS-3,4-ON): β-glutathionyl-α-(3,4-dimethoxyphenyl)-ethan-1-on].

To produce GSH-conjugates, prior enzymatic reactions with β-etherases had to be accomplished. The substrate *rac*-**11** was prepared from **1** by an enzymatic reaction with β-etherases LigE and LigF-NA. Similarly, **1** was incubated separately with β-etherases LigE or LigF-NA to synthesize the two enantiomers (*S*)-**11** and (*R*)-**11**, respectively. The achiral model substrate **12** was synthetized using β-etherase LigF-NA as a biocatalyst and **2** as the substrate. All subsequent glutathione lyase reactions were analyzed by HPLC (**Tab. 3.4**).

Tab. 3.4: Specific activities of glutathione lyases determined in reactions with different model substrates.

Enzyme	Activity towards *rac*-**11** (U/mg)	Activity towards (*R*)-**11** (U/mg)	Activity towards (*S*)-**11** (U/mg)	Rate (*R*)/ Rate (*S*)	Activity towards **12** (U/mg)
LigG	5.01	5.42	0.62	9	3.6
LigG-NS	14.8	16.9	0.42	42	2.8
LigG-TD	11.0	13.3	6.21	2	2.0

LigG-NS displayed the highest specific activity towards racemic **11**. LigG and LigG-NS show strong preference for conversion of the (*R*)-substrate over the (*S)*-enantiomer. In contrast, LigG-TD converts (*R*)-**11** only twice as fast as (*S*)-**11**, which makes it the least stereoselective glutathione lyase among the tested enzymes. Specific activities toward achiral substrate **12** show relatively different trend. LigG exhibit the highest value of 3.6 U/mg, while the values of specific activities of LigG-NS and LigG-TD are ca. 20 % and ca. 45 % lower, respectively. Compared to the specific activities towards substrate **11**, it can be deducted that the tested glutathione lyases show significant structural preference towards chiral substrate.

3.2 Whole-cell catalysis using β-etherases and glutathione lyases

One of the major limitations of β-etherases and glutathione lyases for biocatalytic applications is their dependence on GSH. This tripeptide, naturally occurring in most living organisms, can exist in a reduced (GSH) and oxidized (GSSG) form. GSSG is converted back into two molecules of GSH by glutathione reductase, an enzyme commonly present in living cells. The use of glutathione reductase from *Allochromatium vinosum* for GSH recycling in an enzymatic cascade employing β-etherase and glutathione lyase for lignin depolymerization was already tested by Reiter *et al*. in 2013 [119]. Considering the significant amounts of GSH present in *E. coli* cells together with the enzyme glutathione reductase, the use of a whole-cell catalyst combining the most effective β-etherases and glutathione lyase in *E. coli* was attempted for intracellular β-*O*-4 aryl ether cleavage.

3.2.1 Cloning of β-etherases and glutathione lyase

Based on the information gained from experiments earlier, it was decided to combine the two most active and enantiocomplementary β-etherases LigE and LigF-NA as well as the least enantioselective glutathione lyase LigG-TD in one whole-cell catalyst. The pETDuet-1 vector was selected as the expression vector for β-etherase genes as it harbours two multiple cloning sites (MSC) under control of the T7 promoter. Both genes were available in vector pET28a(+) [110]. Hence, the NdeI restriction site of *ligE* was replaced by NcoI for subsequent cloning into MSC-1 of pETDuet-1 using restriction sites NcoI and HindIII. Successful insertion of the gene was verified by colony PCR and sequencing, yielding vector pETDuet1_*ligE*. In the second step, the *ligF-NA* gene was cloned to MCS-2 of pETDuet1_*ligE* using restriction sites NdeI and XhoI. The resulting vector pETDuet1_*ligE_ligF-NA* was again confirmed by colony PCR and sequencing and is shown in chapter 7.4. At the same time, the gene encoding LigG-TD was cloned from pET28(+)_*ligG-TD* into vector pIT2 using NdeI and HindIII restriction sites. Vector pIT2 [132] was chosen due to its compatibility with pETDuet-1 (different origins of replication and different selection markers; both inducible by IPTG). The resulting vector pIT2_*ligG-TD* was confirmed by colony PCR and sequencing and is shown in chapter 7.4.

3.2.2 Co-expression of both β-etherases and the glutathione lyase

Two different expression hosts, *E. coli* BL21 (DE3) and *E. coli* C43 (DE3), were co-transformed with pETDuet1_*ligE_ligF-NA* and pIT2_*ligG-TD*. The presence of both

plasmids in the strains was ensured by growth on solid media containing ampicillin (for pETDuet-1 selection) and tetracycline (for pIT2 selection). Heterologous co-expression of LigE, LigF-NA and LigG-TD was tested to achieve the best expression levels, and at the same time provide the highest enzyme activity. The initial expression experiment using *E. coli* strains BL21 (DE3) and C43 (DE3) was performed at 20˚C using Terrific Broth (TB) media. The activity of the produced whole-cell biocatalyst was concurrently tested in the conversion of **1**. HPLC analysis confirmed the formation of 2,6-MP, GS-βVG and the final product VG in both *E. coli* strains. The reaction using *E. coli* C43 (DE3) led to a complete conversion of substrate **1** within 5 h of reaction along with quantitative formation of 2,6-MP and VG. In contrast, the *E. coli* BL21 (DE3) strain did not achieve full conversion in either of the two steps of the cascade reaction, even after 20 h (1320 min) (**Fig. 3.5**). Hence, *E. coli* C43 (DE3) seemed to be the better host for co-expression of the target proteins and, thus, was selected as host for all further experiments.

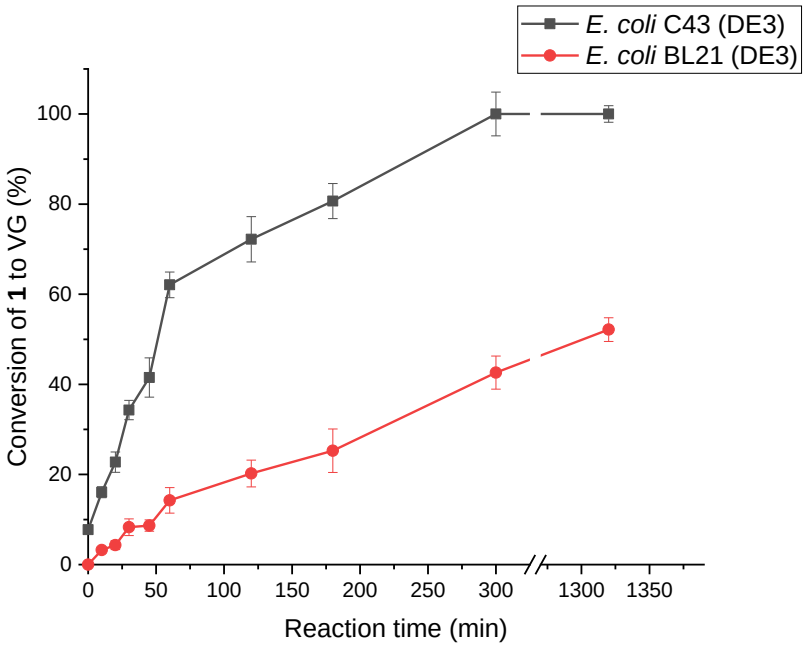


Fig. 3.5: Conversion of substrate **1** using two different strains of whole-cell catalyst (4.33 g/L dry cell weight in 1 mL reaction). Each reaction was carried out in triplicate using 0.2 mM of substrate.

The following expression experiments were carried out at 17, 20, 25, 30 and 37°C and the possible impact of the length of expression was tested after 18 h and 44 h. The goal was to achieve the highest activity of the whole-cell catalyst. For this, the activity of the resulting *E. coli* C43 (DE3) cells harboring LigE, LigF-NA and LigG-TD upon expression at different temperatures and for different cultivation times was compared. As a result, the highest activity towards **1** was obtained after expression at 20°C for 18 h (**Tab. 3.5**).

Higher or lower expression temperatures, as well as a prolonged cultivation time resulted in an evident decrease in activity.

Tab. 3.5: Activities of whole-cell catalyst *E. coli* C43 (DE3) (pETDuet1_*ligE*_ *ligF*-NA) (pIT2_*ligG*-TD) towards the model substrate **1** depending on the length and temperature of expression.

Length of expression (h)	18					44			
Temperature of expression (˚C)	17	20	25	30	37	20	25	30	37
Activity (mU/g_{DCW})	55	81	44	18	17	37	30	18	9

SDS-PAGE was used in an attempt to confirm expression of the three proteins. This, however, was complicated due to the very similar sizes of the proteins: 34 kDa, 31 kDa and 33 kDa for LigE, LigF-NA, LigG-TD, respectively. In order to improve the separation of the proteins on an SDS gel, the acrylamide concentration in the separation gel was increased to 20 % (compared to the standard 12 %), unfortunately with no enhancement in protein separation (**Fig. 3.6**).

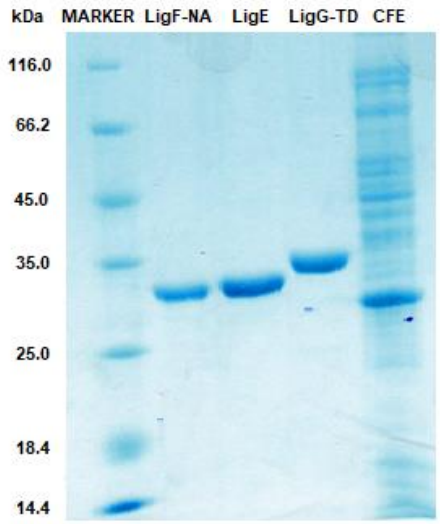


Fig. 3.6: SDS-PAGE analysis of purified enzymes LigF-NA, LigE and LigG-TD as well as cell-free extract (CFE) of *E. coli* C43 (DE3) harboring LigE, LigF-NA and LigG-TD. Marker: PageRuler Prestained Ladder (Life Technologies).

Despite the inefficient separation of the proteins, sufficient expression of each protein was confirmed by chiral and achiral HPLC assay, as well as proteomic analysis of the produced cells. Proteomic investigation affirmed major amounts of LigE compared to Lig-FA and LigG-TD, which in our whole-cell reactions presents an advantage since LigE shows lower activity towards lignin model **1** compared to LigF-NA.

3.2.3 Determination of pH optima

Expressed β-etherases and glutathione lyase possess pH optima in the alkaline range (9-9.5) [105,110], which allows a nucleophilic attack of the glutathione thiolate on the β carbon of the substrate. The whole-cell approach, however, has to consider conditions necessary for the bacterial metabolism since the desired glutathione regeneration should occur in the cytosol. The pH optimum of the whole-cell reaction using **1** as the substrate was analyzed to achieve highest activity of the whole-cell biocatalyst (**Fig. 3.7**). Here, the whole-cell reaction displayed an optimum in conversion at pH 8.0.

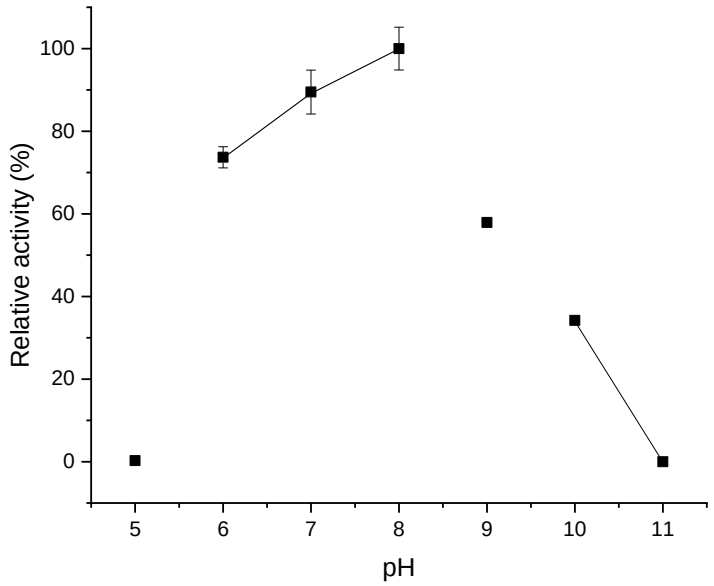


Fig. 3.7: pH profile of *E. coli* C43 (DE3) (pETDuet_*ligE*_ *ligF-NA*) (pIT2_*ligG-TD*)-catalyzed conversion of **1** [pH 5: 100 mM sodium acetate buffer, pH 6: 100 mM potassium phosphate buffer, pH 7: 100 mM potassium phosphate buffer, pH 8: 100 mM potassium phosphate buffer, pH 9: 100 mM glycine buffer, pH 10: 100 mM carbonate buffer, pH 11: 100 mM carbonate buffer]. Reactions were performed in triplicate. Values are given as relative activities (%) with the highest measured activity (81 mU/g_{DCW}) set to 100 %.

As expected, the pH optimum of the whole-cell biocatalyst does not copy the pH optimum of the expressed proteins, but ensures optimal conditions required by the cells to perform an active cell metabolism for provision and recycling of GSH.

3.2.4 Impact of co-solvent and substrate concentration

Due to the low solubility of substrate **1** in aqueous solution, it has to be dissolved in an organic solvent for the cascade reaction. So far, only DMSO had been tested as a possible co-solvent when using isolated β-etherases and glutathione lyases [105,110]. For larger-scale applications, however, DMSO would not be the solvent of choice due to its

low volatility and resulting problems during product work-up. Hence, besides DMSO, also the polar co-solvents methanol and isopropyl alcohol were selected for their ability to efficiently dissolve the model substrate, and their impact on the whole-cell activity at different solvent concentrations (0.5–12.5 % v/v) was studied. Results revealed that at low solvent concentration (between 0.5 % and 5.0 % v/v), the use of isopropyl alcohol or methanol as co-solvents yielded higher activity of the whole-cell biocatalyst compared to DMSO. In contrast, at higher solvent concentrations (≥5.0 % v/v), only methanol still appeared to be a good alternative to DMSO, as here isopropyl alcohol resulted in a significant activity decrease (**Fig. 3.8**).

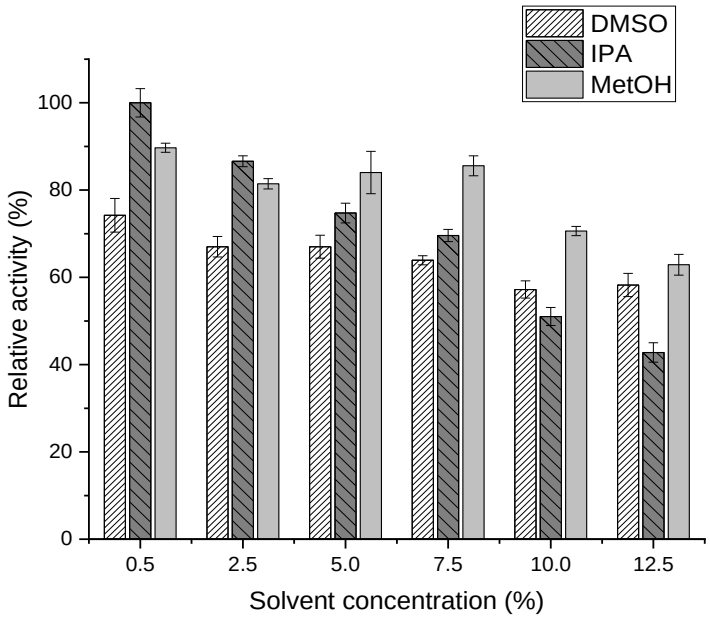


Fig. 3.8: Impact of co-solvent concentration on the whole-cell catalyst activity. The concentration of **1** was kept at 0.2 mM. Values are given as relative activities (%) with the highest measured activity (243 mU/g_{DCW}) set to 100 %.

Aliphatic alcohols have been reported to damage cellular membranes [133]. Alcohols with a longer carbon chain length are more destructive than short ones. Hence, methanol is a better co-solvent choice for use in whole-cell conversions at higher concentrations than isopropyl alcohol. For comparison, reactions with 0.2 and 1 mM of **1** were performed without co-solvent addition. In these, specific activities decreased by approx. 20 % for the reaction with 0.2 mM of **1** and by ca. 70 % for the reaction with 1 mM of **1** compared to reactions using the same substrate concentrations with the addition of 5 % v/v organic solvent (**Fig. 3.9**).

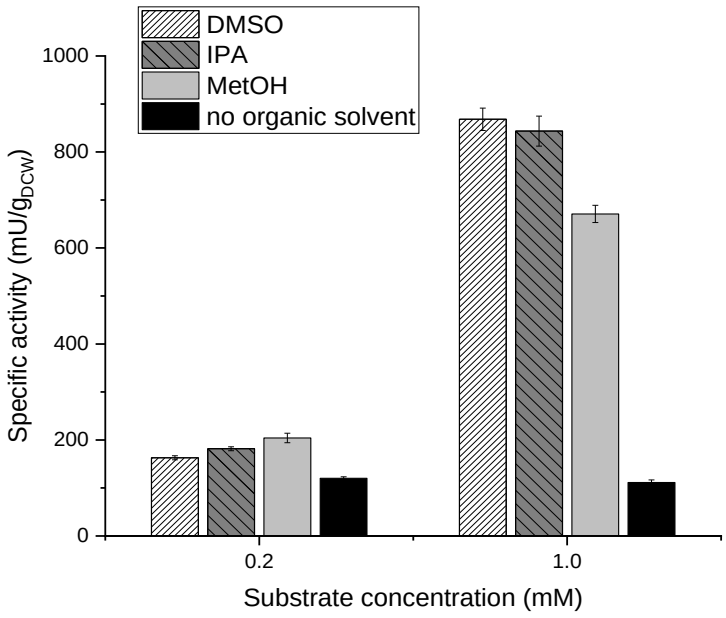


Fig. 3.9: Impact of co-solvent addition on the specific activity of *E. coli* C43 (DE3) (pETDuet_*ligE*_ *ligF-NA*) (pIT2_*ligG-TD*) cells towards **1**. In case of co-solvent addition, the organic solvent concentration was kept at 5.0 % v/v. All reactions were performed in triplicate.

This confirms the positive effect of co-solvent addition on the reaction rate of our whole-cell conversions by solubilizing the substrate and probably also permeabilizing the cells for efficient substrate uptake.

In addition, the impact of increasing substrate concentration, dissolved in the three previously tested co-solvents, was investigated. Thus, analytical-scale conversions were performed with substrate concentrations ranging from 0.2 to 10 mM, while the co-solvent concentration in the reaction mixture was kept at 5.0 % v/v. Subsequent HPLC analysis revealed that DMSO and methanol were superior compared to isopropyl alcohol at substrate concentrations ≥1.5 mM (**Fig. 3.10**).

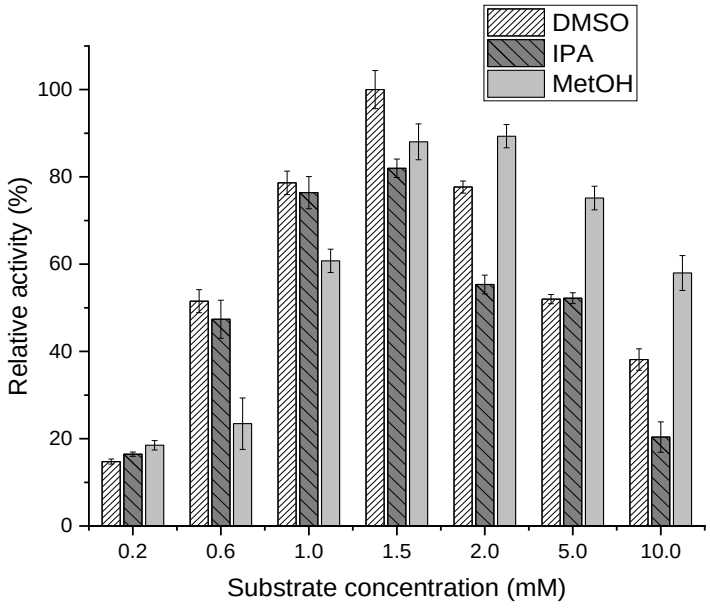


Fig. 3.10: Impact of substrate **1** concentration on the whole-cell activity. Co-solvent concentration was kept at 5.0 % v/v. Values are given as relative activities (%) with the highest measured activity (1.1 U/g_{DCW}) set to 100 %. All reactions were performed in triplicate.

With 5.0 % v/v isopropyl alcohol as a co-solvent, a significant drop in activity was observed for substrate concentrations above 1.5 mM. This is to a lesser extent also observed with DMSO as the co-solvent. In comparison, methanol exhibited the lowest negative impact on the whole-cell activity in reactions with >2.0 mM substrate. Based on these results, methanol seems to be the best co-solvent choice among the tested ones to be applied in semi-preparative-scale conversions of our lignin model substrate, where a high substrate concentration is desired.

3.2.5 Impact of glucose addition

Depletion of GSH within the cell causes an activation of the pentose phosphate pathway, which produces NADPH and different sugar phosphates starting from glucose. The generated sugars can serve as precursors for the de novo synthesis of glutathione, while the NADPH cofactor is required by glutathione reductase to convert GSSG into GSH [134]. For this reason, the addition of glucose to the whole-cell conversion was performed to supply the cells with a carbon and energy source for efficient NADPH and GSH production. Initially, the whole-cell reaction was performed with 0.8 mM of model substrate **1** and addition of 10 mM glucose. Interestingly, no improvement in conversion was observed upon glucose addition (**Fig. 3.11**). To check whether glucose addition might be beneficial at higher substrate concentrations, the whole-cell reaction was repeated using 10 mM of substrate **1**. This substrate concentration should exceed the

expected intracellular GSH concentration already by a factor of 57 or more [135,136]. Nevertheless, glucose addition did not result in higher conversion of the lignin model substrate compared to reactions without glucose addition (**Fig. 3.11**). In both cases, however, no full substrate conversion was achieved within 22 h of reaction (48 % conversion with and 49 % conversion without glucose addition), suggesting that the limitation in the reactions is not caused by the lack of carbon source (GLC) in the reaction, but may be caused by slow reaction rate causing the degradation of the cells or by substrate inhibition.

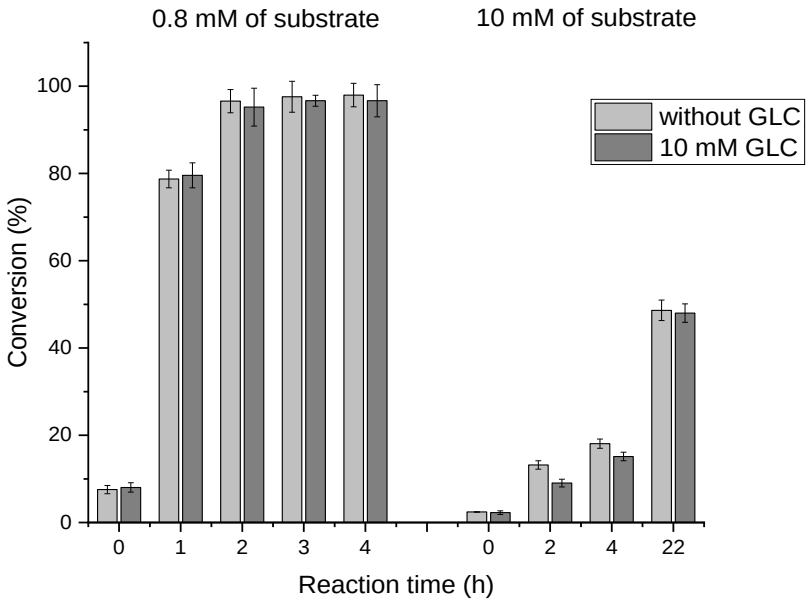


Fig. 3.11: Impact of glucose (GLC) addition on the conversion of **1** to final product VG at two different substrate concentrations. Reactions were performed in triplicate.

A higher solvent (DMSO) concentration (15 % v/v and 25 % v/v) was tested in the reaction with 10 mM of **1** to investigate the possible impact on the conversion rate. Increased solvent concentration ensures larger amounts of substrate to be dissolved and thus more substrate could be converted. However, as shown before, the increased organic solvent concentration has damaging effect on the biocatalyst. Investigation of the impact of 10 mM glucose addition to the reactions was also tested in these “harsher” conditions. HPLC analysis allows simultaneous analysis of both reaction steps of the cascade, and thus enables calculation of conversions for every step of the cascade separately. The result of the conversion after 22 h reactions are summarized in **Tab. 3.6**.

Tab. 3.6: Conversion of substrate **1** to intermediate *rac*-**11** (first step of the cascade, catalyzed by β-etherases), conversion of *rac*-**11** to product VG (second step of the cascade, catalyzed by glutathione lyase) and overall conversion of **1** to final product VG by *E. coli* C43 (DE3) (pETDuet1_*ligE_ligF-NA*) (pIT2_l*igG-TD*) at increased DMSO concentrations compared to the standard reaction with 5 % v/v of DMSO with and without glucose addition.

Concentration of DMSO (%)	Glucose addition	Conversion of **1** to *rac*-**11** (%)	Conversion of *rac*-**11** to VG (%)	Overall conversion of **1** to VG
25	-	18.3	65.2	11.9
25	10 mM	23.5	94.2	22.1
15	-	23.0	80.8	18.6
15	10 mM	27.0	96.3	26.0
5	-	48.9	99.9	48.9
5	10 mM	48.2	99.9	48.2

Based on the HPLC data, the increased amount of DMSO had a negative effect on the conversion rate of both steps of the cascade. Interestingly, the increased solvent concentration had a larger effect on the reaction catalyzed by β-etherases (first step of the cascade) than on the reaction catalyzed by glutathione lyase (second step of the cascade), especially when glucose was added to the reaction. The low conversion rate of the first reaction could be caused by limited amount of LigF enzyme available since proteomic analysis confirmed mostly LigE expression. In that case, mainly one enantiomeric form of substrate would be converted, creating a “bottleneck:” of the cascade. Additionally, results show that when higher amount of the organic solvent in the reaction is used, the addition of the glucose has beneficial impact on the conversion. In the reaction with 25 % of DMSO, the first step of the reaction exhibit improvement in conversion by approx. 5 % and the second by ca. 30 % when comparing to the reaction without glucose addition. In the reaction with 15 % of DMSO, the addition of glucose enhanced the conversion of the first and second step of the conversion by approx. 4 % and 16 %, respectively. Beneficial impact of glucose addition in conversion rate in higher DMSO concentration is probably caused by better stress resistance. Cells with carbon source (GLC) available, can use the glucose for biosynthetic or assimilatory reactions, that are necessary due to the damage caused by high DMSO concentration (physical damage, cell wall disruption). Despite the fact that positive impact of the GLC was observed in both steps of the cascade with increased solvent concentration, the overall conversions are at levels of 22.1 % for reaction with 25 % of DMSO and 26.0 % for the reaction with 15 % of DMSO. These conversion rates were significantly lower than in standard reaction with only 5 % of DMSO.

3.2.6 Repeated cell use

One of the major benefits of using a whole-cell catalyst compared to a free enzyme is the possibility to recycle the cells in multiple rounds of reaction. However, drawbacks connected with the repeated use of catalyst include relatively complicated downstream

purification, lost of catalytic activity and viability of the catalyst, and production of unwanted by-products and/or degradation compounds. Due to the fact that the used lignin-model compound **1** is relatively small and hydrophobic and thus is able to permeate the cell wall, the possibility of a repeated use of the whole-cell catalyst in the reaction was explored. This way, the ability of the cells to regenerate the cofactor GSH in multiple rounds of the reaction was analyzed. At the same time, the addition of 10 mM glucose was also tested to inspect its effect on GSH regeneration and the reusability of the whole-cell biocatalyst. All three previously tested organic solvents were applied in reactions using 0.8 mM of **1** for comparison. Reactions were performed in 5 mL scale for 4 hours. Afterwards, the sample was taken and the remaining reaction mixture was centrifuged at 6,000 *g* for 5 min. The supernatant was discarded and the cells were resuspended in the fresh reaction buffer containing substrate. The cell recycle was repeated for four cycles and the initial activities and conversion rates of each 4-hour cycle are shown in **Fig. 3.12**.

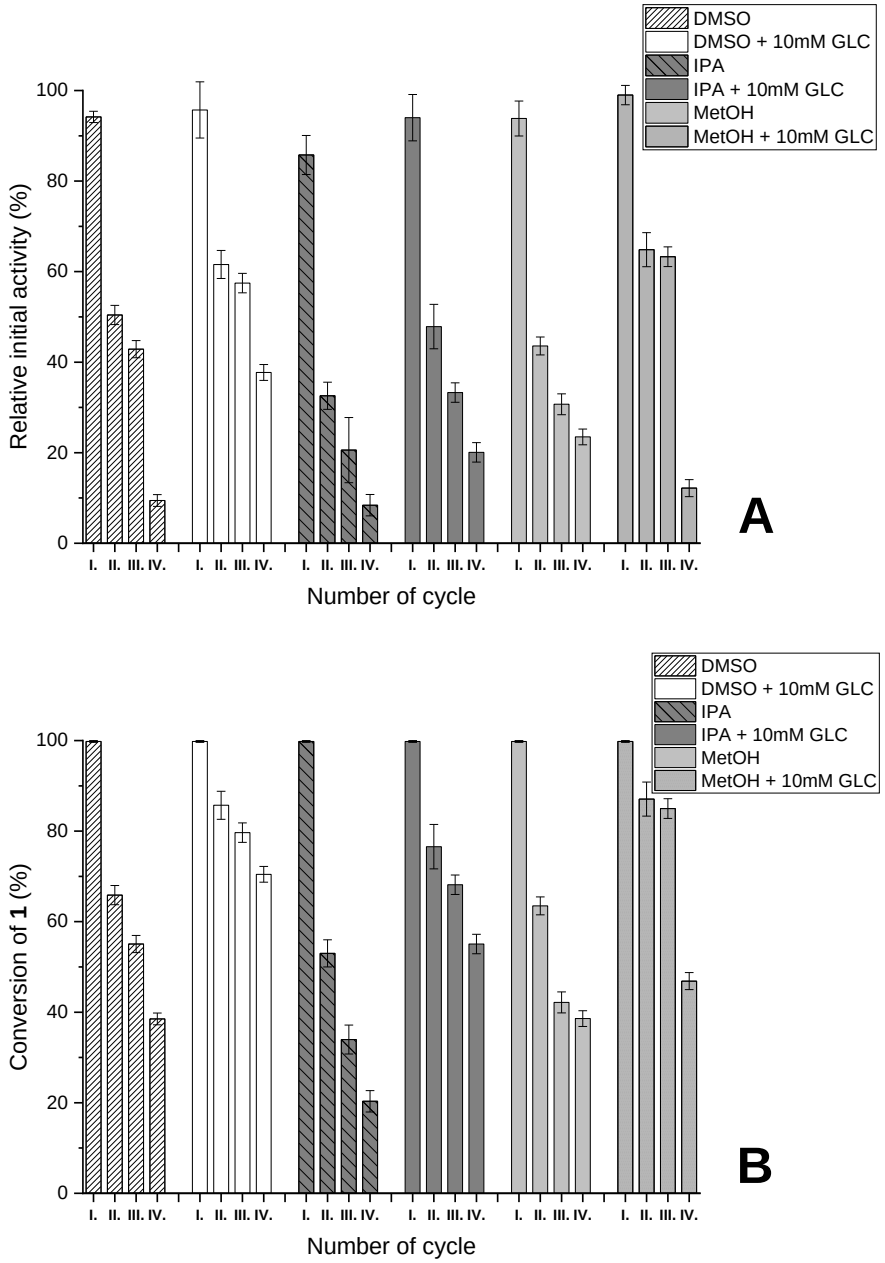


Fig. 3.12: (**A**) Relative initial activity of the whole-cell catalyst towards model substrate **1** in repeated reaction cycles using different solvents with and without the addition of glucose. Values are given as relative initial activities (%) with the highest measured activity (0.73 U/g_{DCW}) set to 100 %. All reactions were performed in triplicate. (**B**) Conversion of the whole-cell catalyst towards model substrate **1** in repeated 4-hour reaction cycles using different solvents with and without the addition of glucose.

Subsequent HPLC analysis revealed several conclusions. As shown before (see section 3.2.5), the impact of glucose addition on the initial activity of the catalyst in the first cycle is marginal and independent from the solvent used to dissolve the model substrate. In all cases though, the reactions with added glucose exhibit slightly higher initial relative activity (≤6 %). In agreement with previous results, using DMSO and MeOH as a co-solvent result in a better reaction performance compared to isopropyl alcohol. Moreover,

for the second and third cycles of the reaction with added glucose, higher initial relative activities are obtained compared to their glucose free counterparts. In the second round of reactions using DMSO, IPA and MeOH with glucose addition, 11 %, 15 % and 21 % improvements in the initial activity, respectively, compared to the reactions with no glucose are observed. In the third round of the reactions, this improvement due to glucose addition is even higher: 14 %, 12 % and 32 % for DMSO, IPA and MeOH, respectively, compared to the respective reactions with no added glucose. For the fourth cycle of the reaction with added glucose, higher initial relative activity was observed with solvent DMSO and IPA by 28 % and 12 % respectively when compared to the reactions without glucose. Surprisingly, the fourth cycle of the reaction using MeOH as a solvent is the only case when relative activity was higher in glucose-free reaction by 11 %. Overall, glucose appears to have a beneficial impact on the initial activity during the repeated use of the whole-cell catalyst. Glucose addition might strengthen the cells towards the mechanical stress they are exposed to (centrifugation, resuspending) and also prevent cell lysis, which was observed in reactions without GLC, by providing a preferred carbon source.

When considering the overall yield of the reactions with the reused biocatalyst regarding the production of the final product VG after four cycles of each 4-hour reactions, the preparation with DMSO+GLC reached the best yield of 41.7 mM_{VG}/g_{DCW}, followed by the IPA+GLC reaction with 39.6 mM_{VG}/g_{DCW} and MeOH+GLC with 39.0 mM_{VG}/g_{DCW}. Reactions without GLC using DMSO, MeOH and IPA achieved yields of 34.0 mM_{VG}/g_{DCW}, 29.8 mM_{VG}/g_{DCW} and 26.5 $mM_{VG}/g_{DCW,}$ respectively. Based on the results, it can be concluded that, overall, DMSO had the lowest negative impact on the whole-cell biocatalyst, probably due to the fact that alcohols are much better in permeabilizing the cell membrane than DMSO [133,137]. Therefore, higher DMSO concentrations are usually compatible with living cells whereas alcohols are inhibiting cell growth already at much lower concentration [138]. Regardless of the improved yield and conversion levels of the cascade with glucose addition, the observed decrease of conversion in every consecutive reaction cycle with the reused catalyst is not in accordance with an economically viable process. In order to improve the economy of the cascade, the application of immobilized whole cells in repeated use as well as a substrate feeding process instead of catalyst reuse could be tested in the future.

3.2.7 Kinetic resolution of *rac*-2,6-MP-VG (1) using the whole-cell catalyst

Since β-etherases are highly stereoselective enzymes, our *E. coli* whole-cell approach can be applied in the kinetic resolution of lignin model substrates if only one β-etherase is present. This way, enantiopure substrate and enantiopure glutathione adduct can be accessed. As proof of concept, *E. coli* cells harboring either LigE together with LigG-TD or LigE alone as well as cells containing either LigF-NA together with LigG-TD or LigF-NA alone were employed in reactions with **1**. Application of (*R*)-selective β-etherase LigE will yield (*S*)-2,6-MP-VG and GS-β(*S*)VG, whereas (*S*)-selective LigF-NA will produce (*R*)-2,6-MP-VG and GS-β(*R*)VG. If the glutathione lyase LigG-TD is present, the glutathione adducts GS-β(*S*)VG and GS-β(*R*)VG will be further converted to GSSG and VG and hence, only enantiopure substrate remains. Initial kinetic resolution experiments were performed in a final volume of 1 mL and with 10 mM substrate. The four *E. coli* strains *E. coli* C43 (DE3)(pET28a(+)*_ligE*)(pIT2*_ligG*-TD) and *E. coli* C43 (DE3) (pET28a(+)*_ligF*-NA) (pIT2*_ligG*-TD) as well as *E. coli* C43 (DE3) (pET28a(+)*_ligE*) and *E. coli* C43 (DE3) (pET28a(+)*_ligF*-NA) were used in parallel and results were compared (**Tab. 3.7**).

Tab. 3.7: Enantioselectivity E of whole-cell reactions in the conversion of **1** (*rac*-2,6-MP-VG) as determined by chiral HPLC [ee_S: enantiomeric excess of substrate].

Reaction time (h)		30			48	
Strain	ee_S (%)	Conversion (%)	E-value[a]	ee_S (%)	Conversion (%)	E-value[a]
E. coli C43 (DE3) *ligE*	21.8	18	>100	35.4	27	>100
E. coli C43 (DE3) *ligE_ligG-TD*	99.9	50	>100	99.9	50	>100
E. coli C43 (DE3) *ligF-NA*	69.7	42	>100	89.4	48	>100
E. coli C43 (DE3) *ligF-NA_ligG-TD*	99.9	50	>100	99.9	50	>100

[a]Calculated according to Chen *et al*. [130].

E. coli cells not harboring the glutathione lyase turned out to be less efficient in the kinetic resolution of **1**, as even after 48 h of reaction no complete separation of substrate enantiomers was observed. In contrast, cells also containing LigG-TD displayed 50 % conversion after 30 h and the chiral HPLC chromatograms confirmed formation of enantiopure compounds. The significantly higher efficiency in reactions with glutathione lyase (resulting in a higher enantiopurity of leftover substrate) compared to reactions without LigG-TD confirms that the glutathione lyase step, and hence glutathione recycling within the cell, is required for efficient conversion of 5 mM substrate (corresponding to the amount of one enantiomer at 10 mM concentration of racemic

substrate). This glutathione recycling is likely performed by glutathione reductase, which is present in the *E. coli* cytoplasm [139]. Since no glucose addition was required to reach full conversion of one substrate enantiomer, the energy status of the cells was obviously sufficient to supply the required NADPH for GSH regeneration. For comparison, whole-cell reactions were also performed by combining *E. coli* cells that separately produced β-etherase (either LigE or LigF-NA) and glutathione lyase LigG-TD. A higher conversion rate of (*R*)-2,6-MP-VG and (*S*)-2,6-MP-VG was obtained using *E. coli* strains where both β-etherase and glutathione lyase were co-expressed (**Fig.3.13**). This further confirms that β-etherase and glutathione lyase have to be present within the same cell to enable efficient GSH recycling.

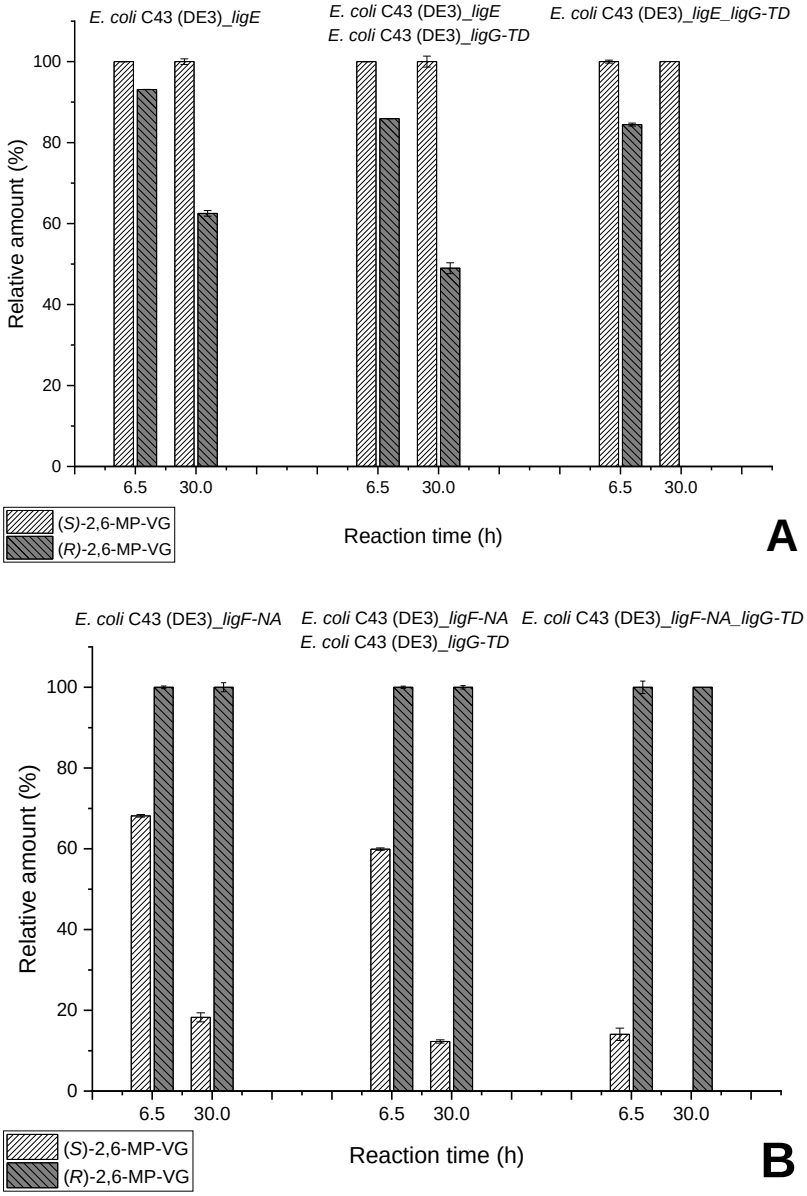


Fig. 3.13: (**A**) Comparison of kinetic resolutions of *rac*-2,6-MP-VG (**1**) using either *E. coli* C43 (DE3) (pET28a(+)*_ligE*), a combination of *E. coli* C43 (DE3) (pET28a(+)*_ligE*) and *E. coli* C43 (DE3) (pET28a(+)*_ligG-TD*), or *E. coli* C43 (DE3) (pET28a(+)*_ligE*) (pIT2*_ligG-TD*) to obtain (*S*)-2,6-MP-VG. (**B**) Comparison of kinetic resolutions of *rac*-2,6-MP-VG (**1**) using either *E. coli* C43 (DE3) (pET28a(+)*_ligF-NA*), a combination of *E. coli* C43 (DE3) (pET28a(+)*_ligF-NA*), and *E. coli* C43 (DE3) (pET28a(+)*_ligG-TD*), or *E. coli* C43 (DE3) (pET28a(+)*_ligF-NA*) (pIT2*_ligG-TD*) to obtain (*R*)-2,6-MP-VG. Values are given as relative amounts (%) of substrate enantiomers with 5 mM set to 100 %. Reactions were performed in triplicate.

3.2.8 Semi-preparative scale kinetic resolution

To demonstrate the applicability of our whole-cell cascade for the preparation of enantiopure (*S*)- and (*R*)-2,6-MP-VG, semi-preparative scale reactions were performed using 10 mM (108 mg) racemic substrate with methanol as co-solvent (final solvent concentration of 5.0 % v/v). Based on initial results, reactions with *E. coli* C43 (DE3) cells harboring LigF-NA and LigG-TD were carried out at an OD_{600} of 40 to yield optically pure

(*R*)-2,6-MP-VG. In contrast, for the preparation of (*S*)-2,6-MP-VG an OD_{600} of 80 of *E. coli* C43 (DE3) (pET28a(+)_*ligE*) (pIT2_*ligG*-TD) had to be used to achieve full conversion of the (*R*)-configured substrate enantiomer. This higher cell density requirement is probably caused by the stereopreference of the glutathione lyase LigG-TD. Despite the fact that this enzyme is able to catalyze the elimination of glutathione from both glutathione adduct enantiomers GS-β(*R*)VG and GS-β(*S*)VG, it is less efficient with the (*S*)-configured enantiomer that is formed by LigE catalysis [105]. This may cause slower GSH recycling in *E. coli* cells harboring LigG-TD together with LigE. After 46 h of reaction, complete depletion of one substrate enantiomer (corresponding to 50.0 % conversion) was achieved in both cases. Reaction work up gave 43 mg of enantiopure (*S*)-2,6-MP-VG and 39 mg of enantiopure (*R*)-2,6-MP-VG (**Fig. 3.14**) in 80 and 72 % isolated yield, respectively. The specific rotation values of both enantiomers were determined to be $[\alpha]_{20}^{D}$ = -21.85° for (*S*)-2,6-MP-VG and $[\alpha]_{20}^{D}$ = +21.15° for (*R*)-2,6-MP-VG.

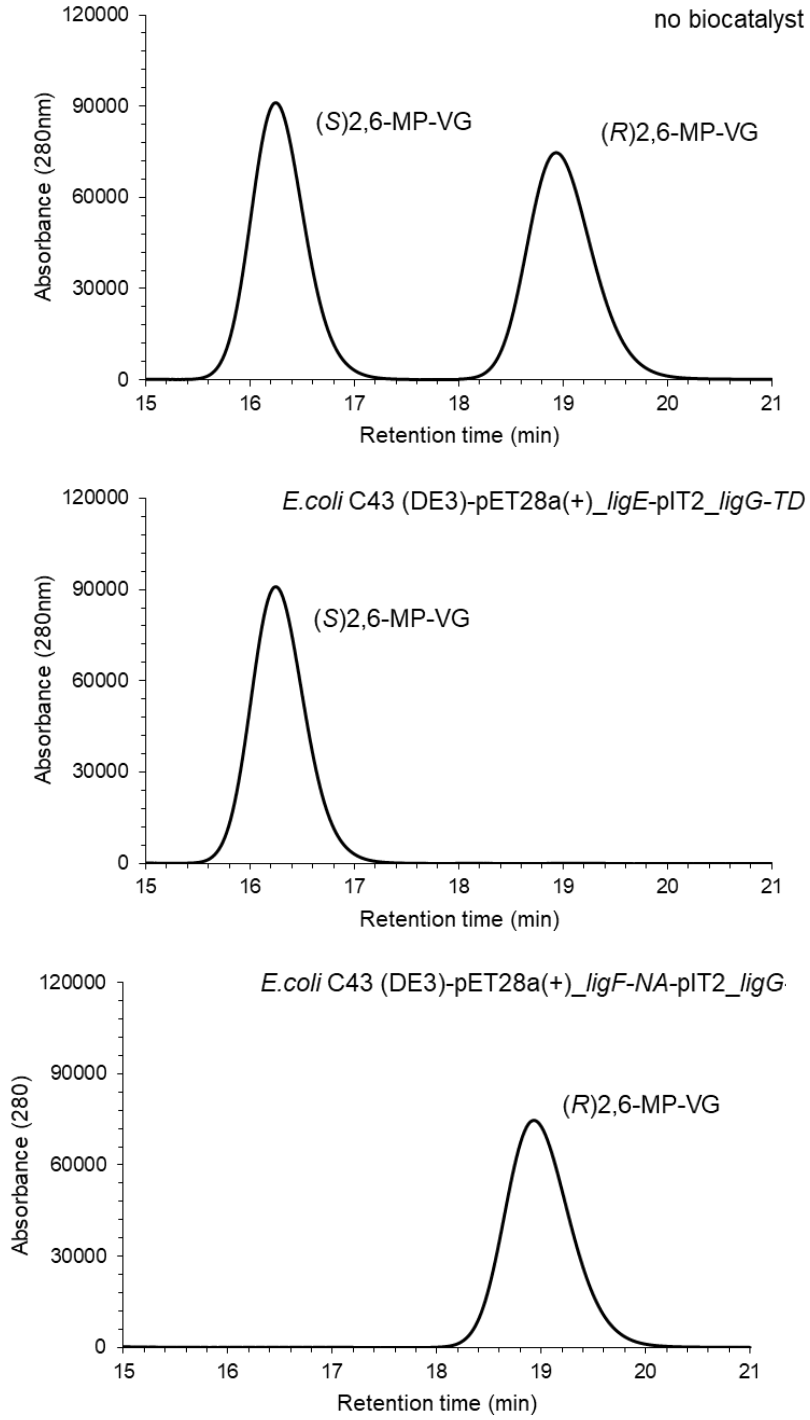


Fig. 3.14: Chiral HPLC chromatograms of *rac*-2,6-MP-VG (top) and pure enantiomers obtained after kinetic resolution of racemic substrate on semi-preparative scale (middle and bottom).

3.3 Mutational analysis of β-ethereases LigE and LigF

The ability of bacterial β-etherases to cleave the β-*O*-4 arylether bonds in lignin in a predictable, defined manner, presents an important feature for possible industrial lignin upgrading. Several *in vitro* systems for the application of β-etherases in the depolymerization of isolated lignin have been described [27,119,120,122]. However, to ensure

an economically profitable lignin valorizing process, not only should catalytic procedures be improved, but the design of more efficient variants of existing enzymes is currently a major objective. Recently, the crystal structures of the well studied β-etherases LigE and LigF from *Sphingobium sp.* SYK-6 were solved [112]. The exact mechanism of these enantiocomplementary enzymes is an enduring knowledge gap, although several assumptions have been made based on the character of the catalyzed reaction. During catalysis, β-etherases form an GSH-conjugate with stereochemical inversion of the chiral center at the β-carbon (from β(*R*)-substrate to β(*S*)-product for LigE and from β(*S*)-substrate to β(*R*)-product for LigF), typical for an S_N2 reaction mechanism [104,112,116]. Understanding the mechanism details of each enzyme together with identifying important residues is a crucial step in comprehending the enzymatic activity.

3.3.1 Alanine scan of LigE residues

The solved crystal structure of C-terminally truncated LigEΔ255 (PDB code: 4YAN) with the glutathione molecule bound in the active site allowed docking of the model substrate β(*R*)-2,6-MP-VG in the hydrophobic binding pocket by Dr. Hauke Voß using the program AutoDock Vina [140] implemented in Yasara. Based on the location of the model substrate in the enzyme structure, several amino acid residues located in instant proximity, as shown in **Fig. 3.15A**, were chosen for further analysis. Amino acid residues located in the substrate binding pocket, namely Tyr23, Trp107, Phe115, Tyr122, Phe142, and Trp197, were subjected to an alanine scan, in order to identify those positions with great influence on the enzyme's catalytic properties. Amino acid residues Tyr23 and Tyr122 were also exchanged by phenylalanine to investigate the importance of the hydroxyl group in tyrosine.

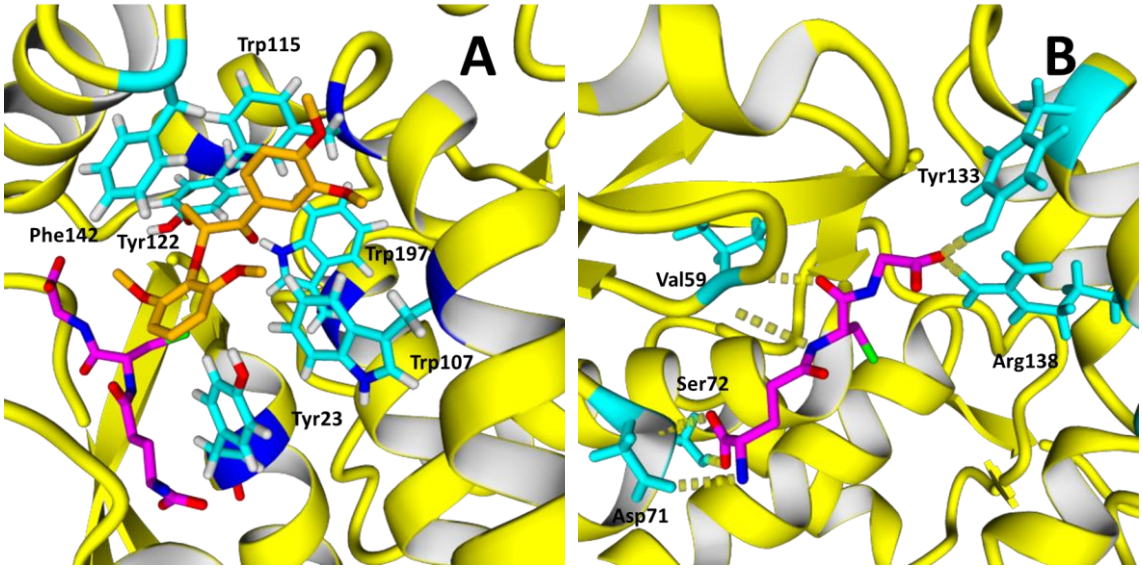


Fig. 3.15: (**A**) Active site of LigE (yellow, PDB code: 4YAN) with the cofactor GSH (magenta) and the docked substrate β(*R*)-2,6-MP-VG (orange). The amino acids in proximity to the docked substrate, chosen to be further analyzed, are Tyr23, Trp107, Trp115, Tyr122, Phe142 and Trp197 (all cyan). (**B**) Amino acid residues of LigE (yellow, PDB code: 4YAN) interacting with GSH (magenta) in its binding pocket: Asp71, Ser72, Val59, Arg138 and Tyr133 (all cyan) with hydrogen bonds shown.

Based on the GSH position within the LigE structure, residue Arg138, which interacts with the glycine moiety of GSH, was chosen to be analyzed (**Fig. 3.15B**). Alanine substitution was performed by site directed mutagenesis using the QuikChange® protocol. Complementary primers were designed using the triplet codons GCU, GCC, GCA, or GCG for alanine and UUU or UUC for phenylalanine. Codons were chosen based on the similarity with the codon encoding the original amino acid residue. Successful mutagenesis was verified by sequencing. The expression strain *E. coli* BL21 (DE3) Gold was transformed with corresponding plasmids pET28a(+) containing the desired mutation, and heterologous expression of the resulting LigE mutants was performed according to chapter 2.2.3.1. Purification was carried out via immobilized metal affinity chromatography (IMAC) using Ni-NTA gravity columns. For all generated LigE mutants, activity towards substrate **1** was determined by HPLC assay and kinetic parameters were obtained by spectrophotometric assay using substrate **3**. Results are summarized in **Tab. 3.8**.

Tab. 3.8: Purification yields (in $mg_{PROTEIN}/L_{MEDIA}$) and relative activities of generated LigE mutants compared to wild-type in conversions of **1** (2,6-MP-VG) and **3** (VN-VG). Values are given as relative activities (%) with the activity of WT LigE towards **1** (1748 mU/mg) and **3** (6876 mU/mg) set to 100 %. Kinetic parameters, K_M and k_{cat}, for LigE and its mutants in the conversion of **3** (VN-VG) were determined at optimal pH and 25˚C. All reactions were performed in triplicate.

Enzyme	Protein yield (mg/L)	Relative activity towards (%)		Kinetic parameters in reactions with **3**			
		1	**3**	K_M (mM)	K_I (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (mM^{-1} s^{-1})
LigE	40	100	100	0.10±0.02	-	6.2±0.15	62
LigE-Y23A	5.5	1.7±0.21	39±2.5	0.43±0.07	4.8±1.43	4.1±0.10	0.1
LigE-W107A	22	7.7±0.98	14±1.9	5.1±1.06	3.6±0.99	14±2.34.	2.7
LigE-F115A	25	6.6±1.02	4.9±0.85	1.6±0.13	-	1.7±0.05	1.1
LigE-Y122A	29	1.0±0.01	4.7±0.36	1.0±0.05	-	1.2±0.00	1.2
LigE-F142A	22	30±1.6	0.5±0.02	3.6±0.34	-	0.4±0.03	0.1
LigE-W197A	23	0.4±0.00	2.1±0.12	1.7±0.38	7.3±2.41	1.0±0.06	0.6
LigE-Y23F	25	0.1±0.00	1.2±0.03	1.5±0.15	-	0.4±0.06	0.3
LigE-Y122F	23	39±2.0	63±4.3	0.11±0.01	-	4.1±0.02	37
LigE-R138A	19	n.d.	n.d.	-	-	-	-

n.d. = not detected

A general lower yield of purified proteins among the mutants compared to WT was observed, which is probably caused by the different, less efficient purification method using gravity-flow columns. This gravity-flow purification made use of Ni Sepharose™ 6 Fast Flow resin loaded in a disposable column allowing as well for protein purification via N-terminal His-tag. Compared to the automated process, however, gravity-flow purification does not permit continuous visualization of the desired protein, thus allowing possible loss of the desired protein during i) loading of the cell-free extract to the column (if the amount of resin is not adequate) or ii) during protein elution since only fractions eluted with the highest imidazole concentration (500 mM) were collected and combined to obtain the purified protein (chapter 2.2.3.2). Despite this, the yield of mutant Y23A is reduced more significantly, suggesting a negative influence of the mutation on the enzyme's expression level.

Moreover, all amino acid exchanges resulted in reduced activity in reactions with both substrates, indicating the importance of those residues for substrate binding and catalysis. Interestingly, the relative activity loss of LigE mutants towards substrates **1** and **3** differs in several cases, especially for mutants Y23A and F142A. The difference between both substrates is the substitution pattern at the aromatic ring next to the ether bond. In the enzyme structure, both positions Tyr23 and Phe142 are located in <5 Å distance to that aromatic ring of the substrate, which could explain the substrate-dependent impact of the mutations on enzyme activity.

Moreover, kinetic studies performed with LigE-Y23A, LigE-W107A and LigE-W197A mutants (shown in chapter 7.8) revealed that the reaction velocity first increased with

increasing substrate concentration until it reached a maximum and, as the substrate concentration increased further, the reaction velocity decreased again. This curve progression is indicative of substrate inhibition. In this specific case, the substrate inhibition is likely caused by the fact that racemic substrate **3** was used. As mentioned before, LigE is strictly stereospecific converting only the (*R*)-enantiomer of the substrate. Based on this knowledge in combination with molecular docking, it is assumed that the non-converted (*S*)-enantiomer of the substrate is able to bind to the active site of these mutants and thus acts as an inhibitor. This would suggest that the residues Tyr23, Trp107 and Trp197 are important for enzyme's enantioselective binding.

This substrate inhibition introduces to the kinetics equation the constant K_I, describing the dissociation constant of the inhibitor-enzyme complex. K_I indicates the inhibitor concentration required to produce half maximum inhibition and thus, describes how potent the inhibition is.

Based on the determined kinetic parameters for conversion of substrate **3** (kinetic analyses are shown in chapter 7.8), most mutations seemed to negatively influence both K_M and k_{cat}. Moreover, variant Y122F displays similar k_{cat} and K_M values as LigE WT, whereas exchange of Tyr122 by alanine resulted in a 10-fold higher K_M value and a 5-fold lower k_{cat} value. This suggests that only the aromatic ring of Tyr122 is important for substrate binding and conversion, while the tyrosine hydroxyl group of the residue 122 is not. Furthermore, exchange of Arg138, the amino acid residue interacting directly with GSH, by alanine caused a complete loss of enzyme activity towards both substrates. Therefore, it can be concluded that Arg138 is crucial for GSH binding and thus LigE catalysis.

3.3.2 Alanine scan of LigF residues

For insights into the catalytic mechanism of β-etherase LigF, the same procedure as for LigE was followed. The model substrate β(*S*)-2,6-MP-VG was docked by Dr. Hauke Voß into the crystal structure of C-terminally truncated LigFΔ242 (PDB code: 4TX0) with the glutathione molecule located in its binding site, using the program AutoDock Vina [140] implemented in Yasara. The active site of LigF has a tunnel-like structure and based on their proximity to the model substrate, several amino acid residues were chosen to be subjected to an alanine scan. Possible direct interactions with the model substrate were predicted for residues Phe7, Asn12, Cys107, Trp108, Val110, Ser111, Trp115, Ile119, Ile122, Pro142, Gln144, Lys147, Trp148 and Ile199 (**Fig. 3.16A**). Analysis of the positioning of GSH within the LigF crystal structure revealed several amino acid residues interacting with the γ-glutamyl, cysteinyl, and glycine moiety of GSH (**Fig. 3.16B**). Among

those, Gln39 was chosen for replacement by alanine as well. Results of the alanine scanning experiment are summarized in **Tab. 3.9**.

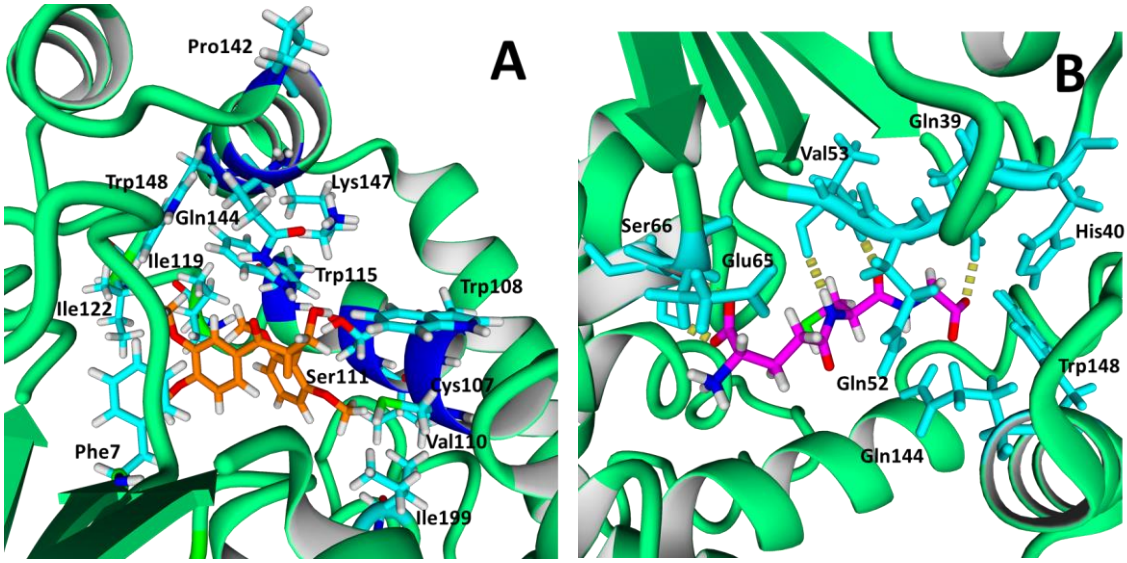


Fig. 3.16: (**A**) Active site of LigF (mint green, PDB code: 4XT0) with the docked substrate β(*S*)-2,6-MP-VG (orange). The amino acids in proximity to the docked substrate, chosen to be further analyzed, are Phe7, Asn12, Cys107, Trp108, Val110, Ser111, Trp115, Ile119, Ile122, Pro142, Gln144, Lys147, Trp148 and Ile199 (all cyan). (**B**) Amino acid residues of LigF (mint green, PDB code: 4XT0) interacting with GSH (magenta) in its binding pocket: Glu65, Ser66, Gln-52, Val-53, Gln-144, His-40, Trp-148, and Gln39 (all cyan) with hydrogen bonds shown.

Tab. 3.9: Purification yields (in $mg_{PROTEIN}/L_{MEDIA}$) and relative activities of generated LigF mutants compared to wild-type in conversions of **1** (2,6-MP-VG) and **3** (VN-VG). Values are given as relative activities (%) with the activity of WT LigF towards **1** (1262 mU/mg) and **3** (2192 mU/mg) set to 100 %. Kinetic parameters, K_M, K_I and k_{cat}, for LigF and its mutants in the conversion of **3** (VN-VG) were determined at optimal pH and 25˚C. All reactions were performed in triplicate.

Enzyme	Protein yield (mg/L)	Relative activity towards (%)		Kinetic parameters in reactions with **3**			
		1	**3**	K_M (mM)	K_I (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (mM^{-1} s^{-1})
LigF	67	100	100	0.93±0.16	3.6±0.97	4.8±0.56	5.3
LigF-F7A	33	0.6±0.12	1.8±0.96	11±8.8	0.18±0.14	0.8±0.59	0.1
LigF-C107A	30	8.2±0.01	9.2±0.63	3.9±2.88	0.83±0.69	1.0±0.63	0.3
LigF-W108A	18	5.9±0.11	5.0±0.15	2.6±1.32	1.4±0.87	0.6±0.27	0.2
LigF-V110A	14	14±1.0	68±1.2	3.3±2.12	0.61±0.46	7.7±0.96	2.4
LigF-S111A	13	0.1±0.00	1.5±0.04	0.74±0.17	3.3±0.84	0.1±0.02	0.1
LigF-W115A	17	n.d.	n.d.	-	-	-	-
LigF-I119A	17	4.0±0.35	3.9±0.12	5.8±0.75	2.2±0.37	0.5±0.01	0.1
LigF-I122A	13	43±1.2	34±1.9	1.2±0.36	1.9±0.73	2.0±0.48	1.7
LigF-P142A	16	149±2.2	184±3.2	1.7±0.76	1.5±0.82	14±4.91	8.4
LigF-Q144A	14	14±0.5	19±2.0	1.5±0.33	2.3±0.69	1.4±0.21	0.9
LigF-K147A	14	0.3±0.03	0.7±0.13	9.1±6.60	0.77±0.61	0.3±0.15	0.0
LigF-W148A	7.6	n.d.	n.d.	-	-	-	-
LigF-I199A	8.3	4.0±0.42	3.9±0.63	1.5±1.46	6.2±1.73	0.7±0.44	0.5
LigF-Q39A	11	n.d.	n.d.	-	-	-	-

n.d.= not detected

As observed for LigE mutants, lower yields of purified LigF mutants were obtained compared to the WT protein (67 $mg_{PROTEIN}/L_{MEDIA}$), probably caused by the less efficient purification method using gravity-flow columns. However, the most significantly affected protein yields were detected in the mutants W148A and I199A (7.6 and 8.3 $mg_{PROTEIN}/L_{MEDIA}$, respectively), indicating a negative impact of the mutations on the enzymes' expression levels.

As observed for LigE mutagenesis, all amino acid substitutions in LigF except P142A led to a reduction in enzyme activity in reactions with both tested substrates. In both cases, the relative activity loss follows a very similar trend. An exception is residue V110A, for which the loss of activity is more severe with substrate **1** than **3**, suggesting a larger impact of this residue on the release of 2,6-dimethoxyphenol from **1** than the release of vanillin from substrate **3**. Determined K_I values show that all mutants except I199A possess a lower K_I than WT (3.6 mM) implicating increased substrate inhibition and hence, a negative effect on specific activity. In case of mutant I199A, however, the K_I value is 1.7-times higher than that of WT, suggesting lower substrate inhibition. On the other hand, the K_M value is 1.7-times lower than that of WT implying a negative effect of mutation I199A on substrate binding and positioning for conversion. Obtained k_{cat} values and catalytic efficiencies of other LigF mutants were generally decreased except for mutant P142A. This LigF mutant displayed significantly improved activity towards both substrates as well as a 3-fold higher k_{cat} value in the conversion of **3**. Hence, this position was further analyzed by site saturation mutagenesis (SSM). In contrast, amino acid substitutions W115A and W148A inactivated the enzyme completely as no product formation in the conversion of substrates **1** and **3** could be observed. To investigate this in more detail, position W115 was also studied further by SSM. Moreover, similar to mutant R138A in LigE, exchange of residue Gln39, interacting with GSH in LigF, by alanine caused a complete loss of enzyme activity. Complete loss of activity in mutants Q39A and W148A (also interacting with GSH [112]) suggests that amino acids interacting with GSH have important role in cofactor binding and/or positioning within the enzyme and thus for the catalysis.

3.3.2.1 Site-saturation mutagenesis of residue P142

Residue P142 of LigF was selected for site-saturation mutagenesis due to the observed improvement in enzyme activity of mutant P142A. For generation of the SSM library of LigF-P142X, primers containing the degenerated NNS codon were designed and the library was produced using the QuikChange® protocol. To investigate the quality of the SSM library, the isolated plasmid library was sequenced. The obtained sequencing result indicated successful randomization of the desired nucleotides (**Fig. 3.17**). In the first and second position of the codon, all fours bases are present in the library, while in the third position only cytosine and guanine are observed, as expected from the used NNS codon for library generation.

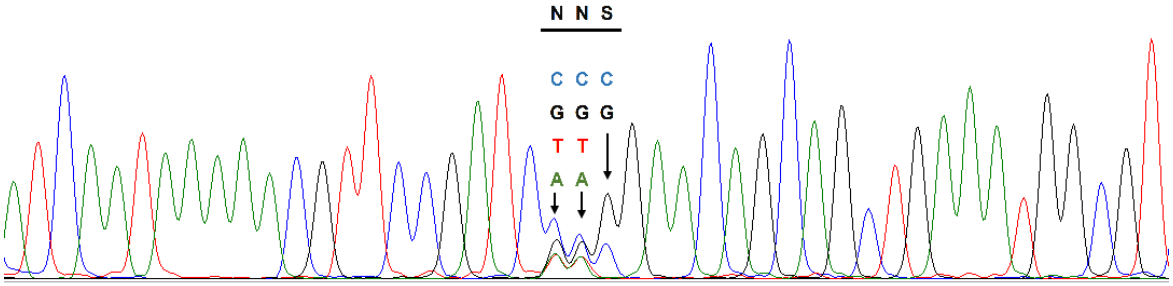


Fig. 3.17: Sequencing result of pET28a(+)ligF-P142NNS with the degenerated NNS codon marked.

After successful transformation of the SSM library, cells were grown and LigF mutants were expressed in a 96-well microtiter plate. Cells were lysed within the microtiter plate using lysozyme or B-PER treatment to obtain cell-free extract. After addition of substrate **3** and the GSH cofactor to each well, the activity of each mutant could be determined spectrophotometrically by measuring the release of vanillin at 360 nm. The measured slopes of vanillin production were first normalized to OD_{600} and the mean value of the wild-type controls. Additionally, the resulting data was converted to relative improvement or deterioration (in percent) of each reaction compared to the mean value of the WT. Final data for the LigF-P142X library in the microtiter plate is shown in **Fig. 3.18**.

	1	2	3	4	5	6	7	8	9	10	11	12
A	**2 (WT)**	**Ø**	-99	-77	-23	-25	-23	93	-17	-20	**Ø**	**-14 (WT)**
B	12	-21	20	27	-5	-20	68	33	-72	-15	-82	-22
C	41	-32	-19	-22	-83	-25	-27	-21	-30	-51	-14	7
D	-49	-18	-47	34	-15	**2 (WT)**	**12 (WT)**	35	-18	-4	31	34
E	-30	-18	-65	-19	18	-28	-11	-2	-30	-97	-13	-1
F	-24	-21	-66	-28	-71	-52	-20	-21	-21	-90	-62	18
G	-23	-20	-21	-19	-21	-22	-93	-16	-88	-30	-8	-87
H	**-8 (WT)**	**Ø**	-19	-35	-21	49	-19	81	-10	22	**Ø**	**5 (WT)**

Fig. 3.18: Results of the activity screening of the LigF-P142X library using the vanillin release assay. Measured slope of the reaction was normalized with OD_{600} and the mean value of the wild type controls and afterwards calculated in % of improvement/deterioration compared to the WT. Wells without cell growth are labeled as **Ø** and wild type controls are marked as **underlined bold**.

Of the screened LigF-P142X library, mutants displaying enhanced activity were chosen to be sequenced and further analyzed. Among these, several hits turned out to be wild type containing different triplets coding for proline. Additionally, several promising non-wild type mutants of LigF-P142X were identified, which were subsequently expressed, purified and characterized. Results are summarized in **Tab. 3.10**.

Tab. 3.10: Purification yields (in $mg_{PROTEIN}/L_{MEDIA}$) and relative activities of promising LigF-P142X mutants compared to wild-type in conversions of **1** (2,6-MP-VG) and **3** (VN-VG). Values are given as relative activities (%) with the activity of WT LigF towards **1** (1262 mU/mg) and **3** (2192 mU/mg) set to 100 %. Kinetic parameters, K_M, K_I and k_{cat}, for LigF and its mutants in the conversion of **3** (VN-VG) were determined at optimal pH and 25˚C. All reactions were performed in triplicate.

Enzyme	Protein yield	Relative activity towards		Kinetic parameters in reactions with **3**			
		1	**3**	K_M	K_I	k_{cat}	k_{cat}/K_M
	(mg/L)	(%)		(mM)	(mM)	(s^{-1})	(mM^{-1} s^{-1})
LigF	67	100	100	0.93±0.16	3.6±0.97	4.8±0.56	5.3
LigF-P142A	16	149±2.2	184±3.2	1.7±0.76	1.5±0.82	14±4.90	8.4
LigF-P142S	43	123±4.8	162±6.8	1.4±0.41	2.6±0.93	8.0±1.61	5.6
LigF-P142Y	68	105±9.2	113±9.3	2.2±1.83	1.5±1.37	10±2.61	4.6
LigF-P142F	16	118±2.9	147±7.4	1.9±0.80	2.1±1.06	7.6±2.36	4.0
LigF-P142I	86	89±3.4	99±2.9	0.78±0.19	4.4±1.43	4.0±0.52	5.1
LigF-P142H	69	134±7.8	116±8.3	1.0±0.27	5.6±2.58	4.0±0.73	4.1

Besides mutant P142I, all identified and characterized LigF-P142X mutants showed improved activity with both substrates compared to the wild type enzyme, with mutant P142A displaying the highest activity increase. Kinetic analyses of these mutants (shown in chapter 7.8) with substrate **3** revealed that this activity increase is caused by significantly higher k_{cat} values, while K_M values are either unchanged or slightly increased. Exceptions are mutants P142I and P142H, for which the obtained kinetic data are similar to wild-type LigF. Moreover, mutants P142A, P142Y and P142F appear to exhibit slightly lower K_I values than the WT, suggesting that despite their increased relative activities, the mutations negatively affected the enzymes' substrate inhibition. Assuming that the inhibition is caused by the non-converted substrate enantiomer, lower K_I values (stronger inhibition) indicate a better binding of (*R*)-**3** which acts as the inhibitor. Interestingly, the relative activity increase was in most cases higher with substrate **3** than substrate **1**, except for mutant P142H. In case of LigF-P142H, activity was only significantly increased with substrate **1**.

3.3.2.2 Site-saturation mutagenesis of residue W115

As substitution of W115 with alanine in LigF resulted in a complete loss of enzyme activity, site-saturation mutagenesis of position W115 was performed with the goal to investigate if tryptophan in this position is crucial for LigF catalysis or if other residues are accepted as well. Complementary primers were designed with altered codons for QuikChange® PCR to enable generation of mutants with all of the 20 standard canonical amino acids at position Trp115. The respective SSM library was successfully cloned, sequenced and mutants were produced in 50 mL media as described in chapter 2.2.3.1. Cell-free extract of each mutant was used to determine activity by HPLC assay using **1** as substrate. Based on this activity test, only three mutants, W115H, W115Y and W115F, were still able to catalyze the cleavage of the β-*O*-4 aryl ether bond in **1**. These mutants were successfully purified and characterized regarding their catalytic parameters, which are summarized in **Tab. 3.11**.

Tab. 3.11: Purification yields (in $mg_{PROTEIN}/L_{MEDIA}$) and relative activities of active LigF-W115X mutants compared to wild-type LigF in conversions of **1** (2,6-MP-VG) and **3** (VN-VG). Values are given as relative activities (%) with the activity of WT LigF towards **1** (1262 mU/mg) and **3** (2192 mU/mg) set to 100%. Kinetic parameters, K_M, K_I and k_{cat}, for LigF and the LigF-W115X mutants in the conversion of **3** (VN-VG) were determined at optimal pH and 25˚C. All reactions were performed in triplicate.

Enzyme	Protein yield (mg/L)	Relative activity towards (%)		Kinetic parameters in reactions with **3**			
		1	**3**	K_M (mM)	K_I (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (mM^{-1} s^{-1})
LigF	67	100	100	0.93±0.16	3.6±0.97	4.8±0.56	5.3
LigF-W115H	29	5.2±0.96	6.5±0.25	0.88±0.16	4.5±1.60	0.1±0.07	0.1
LigF-W115F	25	1.0±0.01	1.5±0.02	4.2±1.66	1.0±0.42	0.4±0.08	<0.1
LigF-W115Y	24	1.0±0.00	0.8±0.00	3.0±1.17	1.5±0.67	0.1±0.04	<0.1

Obtained results indicate that position Trp115 in LigF can be substituted by any other aromatic amino acid in order to keep some enzymatic activity. However, activity values towards substrate **1** and substrate **3** as well as catalytic efficiency are strongly reduced. Moreover, mutants W115F and W115Y exhibit significantly lower K_I values, and hence stronger inhibition, compared to WT as well as higher K_M values, implying that W115 is important for preferential binding of the (*S*)-enantiomer of the substrate (kinetic analyses are shown in chapter 7.8). Based on these results, it is hypothesized that tryptophan at position 115, which is also one of the conserved amino acid residues in LigF homologs, has a large impact on the enzyme's function and that the aromatic character of the residue is important for the direct interaction with the substrate possibly through cation-π interaction. The cation-π interactions are noncovalent binding forces provided by aromatic amino acids, in which the aromatic residues provide a surface of negative electrostatic potential able to bind a wide range of cations [141]. In case of position Trp115, the cation-π interactions could arise between tryptophan or mutants W115H/F/Y (aromatic residues) and Lys147 as a cation, which is located in proximity (<5Å). Cation-π interactions are relatively common in proteins affecting the secondary structure and thus the catalytic properties of the enzymes.

3.3.3 Serine residues involved in LigE and LigF catalysis

Crystal structures of LigE and LigF revealed that in both proteins, serine residues are located in proximity to the thiol of the bound GSH [112]. Several GST classes contain conserved catalytic serine residues, though it has been proven that in some bacterial enzymes, the active site serine was not essential for catalytic activity [142–144]. In case of LigE and LigF, Ser21 and Ser13, respectively, were previously shown not to be essential for direct deprotonation of GSH as mutants S21A and S13A did not cause complete loss of enzyme activity [112]. Nevertheless, both residues may be important for GSH binding in

the active site, orienting the sulfhydryl group of GSH in the catalytic step, or stabilizing the transition state. In that matter, both serine residues were chosen for further analysis. In both cases, NNS libraries were prepared using the QuikChange® protocol and screened in MTP format for activity with the vanillin-releasing spectrophotometric assay. Since the goal was to investigate the influence of this position on substrate binding and conversion, not only improved, but also less active mutants were chosen to be characterized further. In the LigE-S21X library, a high number of wild type enzyme was identified among the active hits, while many mutants completely lost activity. In contrast, the LigF-S13X library contained more active mutants. These were successfully purified and the resulting activity data is summarized in **Tab. 3.12**.

Tab. 3.12: Purification yields (in $mg_{PROTEIN}/L_{MEDIA}$) and relative activities of active LigE-S21X mutants and active LigF-S13X mutants compared to wild-type LigE and LigF in conversions of **1** (2,6-MP-VG) and **3** (VN-VG). Values are given as relative activities (%) with activities of WT enzymes towards **1** and **3** set to 100% [LigE: 1748 mU/mg with **1** and 6876 mU/mg with **3**; LigF: 1262 mU/mg with **1** and 2192 mU/mg with **3**]. Kinetic parameters, K_M, K_I and k_{cat}, for LigE-S21X and LigF-S13X mutants in the conversion of **3** (VN-VG) were determined at optimal pH and 25˚C. All reactions were performed in triplicate.

Enzyme	Protein yield (mg/L)	Relative activity towards (%)		Kinetic parameters in reactions with **3**			
		1	**3**	K_M (μM)	K_I (μM)	k_{cat} (s^{-1})	k_{cat}/K_M (mM^{-1} s^{-1})
LigE	40	100	100	0.10±0.02	-	6.2±0.15	62
LigE-S21A	23	1.4±0.01	3.1±0.01	0.23±0.03	-	0.2±0.00	0.9
LigE-S21G	15	29±1.8	22±0.3	0.48±0.06	-	3.0±0.11	6.2
LigF	67	100	100	0.93±0.16	3.6±0.97	4.8±0.56	5.3
LigF-S13A	26	1.4±0.01	5.4±0.81	1.6±0.42	0.82±0.22	0.6±0.10	0.4
LigF-S13N	28	n.d.	0.2±0.00	7.0±6.51	0.40±0.38	0.1±0.08	<0.1
LigF-S13C	17	12±0.6	68±2.2	1.4±0.42	0.50±0.16	6.1±2.14	4.3
LigF-S13G	17	45±2.5	80±2.3	1.2±0.16	1.6±0.26	7.5±0.70	6.0

All of the tested mutants displayed a (severe) loss of activity in reactions with both substrates, indicating the importance of Ser21 and Ser13 for catalysis of LigE and LigF, respectively. Position Ser21 in the LigE enzyme is predicted to be responsible for activation of GSH in its thiolate form supporting the enzyme's catalytic function. As described by Helmich *et al*. [112], Ser21 seems to be important but not crucial for LigE catalysis, as severe but not complete loss of activity was observed. In the SSM experiment, only one additional amino acid residue, glycine, was accepted at position Ser21 without complete loss of activity. Glycine, as the smallest amino acid with only hydrogen as the side chain, in mutant S21G was able to retain activity towards substrates **1** and **3** in levels of 29 and 22 % compared to WT, respectively. Since glycine has no possibility to form a hydrogen bond via its side chain to support the stabilization of GSH

in its thiolate form, it is hypothesized that another residue is compensating for the absence of the serine-OH. Possibly, Tyr23, which is located in near proximity to the cysteinyl moiety of GSH, could provide the necessary hydrogen bridge. When comparing the specific activities of mutants S21A and S21G, the more dramatic loss of activity in the alanine variant might be explained by the larger size of the residue and possible structural hindering caused by methyl side chain.

Compared to LigE, k_{cat} values of LigE-S21A and LigE-S21G are ~10-fold and ~2-fold lower, respectively, while corresponding K_M values are ~2-fold and ~5-fold higher, indicating that Ser21 is critical for both substrate binding and turnover in LigE.

The SSM experiment performed at position Ser13 of LigF identified multiple residues that retained catalytic activity of the enzyme to some extend. As described before by Helmich *et al*. [112], Ser13 is important but not crucial for LigF catalysis, as the exchange of serine to alanine only led to a serious activity reduction. Our SSM results additionally revealed that glycine, cysteine and asparagine at position Ser13 kept residual catalytic activity of the enzyme, although the activities are evidently lower than for WT. As was suggested in literature, Ser13 together with Asn12 are responsible for GSH activation in LigF [115]. According to the crystal structure of LigF, the conserved Asn12 is located in near proximity (3.258 Å) to the thiol group of GSH and hence, would be able to activate the GSH molecule by stabilizing it in its thiolate form through hydrogen-bonding (**Fig. 3.19**). Glycine at position Ser13 kept activity towards substrate **1** and **3** in levels of 45 and 80 % of WT, respectively. In mutant S13G, the small, uncharged side-chain of glycine probably did not affect the structural properties of the enzyme and enabled Asn12 to activate GSH by deprotonation. In case of mutant S13A, the specific activity towards **1** and **3** declined to 1.4 and 5.4 % of the WT value, respectively, suggesting that the methyl side-chain of alanine affects the enzyme's ability to deprotonate GSH. The question is why in mutant S13A, the expected GSH activation by Asn12 was impaired as well. A possible explanation could be that alanine in position 13 does not only affect the activation of GSH, but also its positioning as well as the positioning of Asn12, and thus could negatively affect the hydrogen bonding between Asn12 and the GSH thiol. In contrast, cysteine at position Ser13 retained 12 and 68 % of activity towards substrates **1** and **3**, respectively. Possibly, the thiol group of cysteine may potentially form a disulfide bridge with the thiol of bound GSH ensuring correct positioning and deprotonation of the GSH within the enzyme. Asparagine in position 13 was found to reduce catalytic activity towards substrate **3** to only 0.2 % of the WT, while the ability to convert substrate **1** was completely lost. Asparagine, like serine, can form a hydrogen bond with the thiol of glutathione, which is important for GSH activation. The size and structural differences of

the residue, however, possibly prevented correct GSH positioning for nucleophilic attack and hence, hindered catalysis. Accordingly, mutants LigF-S13N and LigF-S13A display ~50-fold and ~8-fold lower k_{cat} values, respectively, and ~70-fold and ~16-fold higher K_M values, respectively, than WT LigF. In contrast, mutants LigF-S13C and LigF-S13G, despite the decline in activity levels and an increase of K_M values compared to WT LigF, revealed similar k_{cat} as WT. The observed negative impact on the relative activity of the tested mutants is also influenced by changes in substrate inhibition. All tested mutants displayed significantly lower K_I values compared to WT LigF. While K_I of mutant S13G is two times lower than that of WT, mutants S13A, S13N and S13C displayed even stronger substrate inhibition. These results indicate that upon mutagenesis of Ser13, the affinity for binding of the non-converted (*R*)-enantiomer of the substrate increases, suggesting that possibly Ser13 is important for preferential binding of the right substrate enantiomer as well.

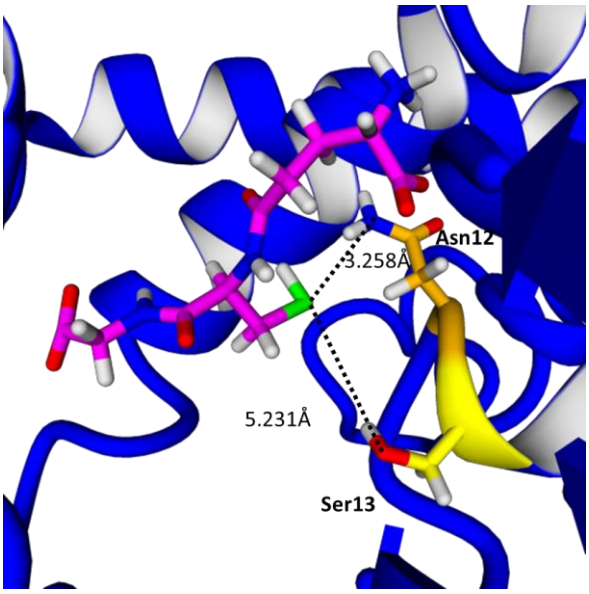


Fig. 3.19: Glutathione binding site (magenta) in LigF (blue, Protein Data Bank code: 4XT0). Residues Ser13 (yellow) and Asn12 (orange) with respective distances to the thiol group of GSH (sulphur atom of GSH highlighted in green) are shown.

Similar to Ser13, the impact of Asn12 on LigF activity was evaluated by site-saturation mutagenesis. Thus, an NNS library was generated at position Asn12 using the QuikChange® protocol. Of the screened LigF-N12X library, mutants displaying improved as well as decreased activity were chosen to be sequenced and further analyzed. All of the generated mutants were expressed, purified and analyzed by HPLC and the spectrophotometric assay. To test the role of Asn12 in combination with the previously

described Ser13, a double mutant was created by substituting both amino acid residues with alanine (LigF-N12A+S13A). **Tab. 3.13** summarizes all of the gathered data.

Tab. 3.13: Purification yields (in $mg_{PROTEIN}/L_{MEDIA}$) and relative activities of active LigF-N12X mutants compared to wild-type LigF in conversions of **1** (2,6-MP-VG) and **3** (VN-VG). Values are given as relative activities (%) with the activity of the WT LigF towards **1** (1262 mU/mg) and **3** (2192 mU/mg) set to 100%. Kinetic parameters, K_M, K_I and k_{cat}, of LigF-N12X mutants in the conversion of **3** (VN-VG) were determined at optimal pH and 25˚C. All reactions were performed in triplicate.

Enzyme	Protein yield (mg/L)	Relative activity towards (%) **1**	**3**	Kinetic parameters in reactions with **3** K_M (mM)	K_I (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (mM^{-1} s^{-1})
LigF	67	100	100	0.9±0.16	3.6±0.97	4.8±0.56	5.3
LigF-N12A	20	2.1±0.03	11±0.9	1.1±0.13	1.4±0.20	0.6±0.07	0.5
LigF-N12H	54	158±2.5	147±1.9	0.8±0.18	5.1±1.61	9.1±1.14	11
LigF-N12M	1.4	84±2.9	76±1.9	0.7±0.19	2.7±0.87	3.5±0.92	5.4
LigF-N12F	22	50±1.7	-	-	-	-	-
LigF-N12R	8.9	0.8±0.02	-	-	-	-	-
LigF-N12T	34	42±2.3	-	-	-	-	-
LigF-N12K	25	5.1±0.31	-	-	-	-	-
LigF-N12A+S13A	28	n.d.	n.d.	-	-	-	-

n.d.= not detected, "-"=not measured

The obtained results revealed that Asn12 is not crucial for catalysis but does have an effect on enzymatic activity. While the K_M values of the tested mutants are rather similar to WT, the K_I value of mutant N12A indicates stronger substrate inhibition compared to WT. In contrast, the only mutant that displayed improved relative activity towards **1** and **3** compared to the WT, LigF-N12H, seems to exhibit slightly reduced substrate inhibition. This together with the significantly improved k_{cat} value are the main driver for the observed higher relative activity of this mutant. Like asparagine at position 12, histidine will also be capable of hydrogen bonding with the $–S^-$ group of GSH through its positively charged imidazole side chain. Moreover, histidine serves as base for deprotonation of serine and cysteine side chains in other enzymes (e.g. proteases or esterases and lipases) due to the suitable pK_a of the imidazole side chain. Hence, in contrast to asparagine, histidine could further support GSH activation by direct deprotonation of the thiol group.

In contrast, mutagenesis of Asn12 and Ser13 together, yielding double mutant LigF-N12A+S13A, led to an inactive enzyme variant, indicating that both residues Asn12 and Ser13 together contribute to GSH activation.

3.4 Substrate binding analyses

Despite the fact that LigF and LigE enzymes display a significant structural difference, they catalyze very similar reactions. Moreover, both enzymes possess rather hydrophobic active sites carrying several tryptophan residues. Aromatic amino acids

(tyrosine, tryptophan, phenylalanine) are responsible for the intrinsic fluorescence of a protein. Tryptophan has a stronger fluorescence, absorption and higher quantum yield than the other aromatic amino acids, and its fluorescence dominates when excited at 280 nm showing a fluorescence peak with maximum at 330–360 nm [145,146]. The quantum yield, intensity and wavelength of the maximum fluorescence is highly sensitive to the microenvironment of the tryptophan molecule [146–148]. Changes of the polarity in the direct surrounding of the molecule affects the fluorescence intensity. Changes in the fluorescence emission spectra of tryptophan-containing proteins can be observed in response to subunit association, denaturation, protein conformational transitions or ligand binding, all of which alter the environment surrounding the indole ring of tryptophan residues. In this work, this property was used in high throughput measurements for investigating the binding of various substrates to the active sites of LigE and LigF. Additionally, several measurements based on isothermal titration calorimetry (ITC) were performed as well.

3.4.1 Substrate binding based on tryptophan fluorescence

3.4.1.1 Substrate binding without GSH addition

To analyze the affinity of the enzyme to bind the substrate, several lignin-model substrates were tested. First, fluorescence emission spectra of LigE and LigF were measured to determine their fluorescence maxima, which were in both cases observed at 340.6 nm, as shown in **Fig. 3.20**.

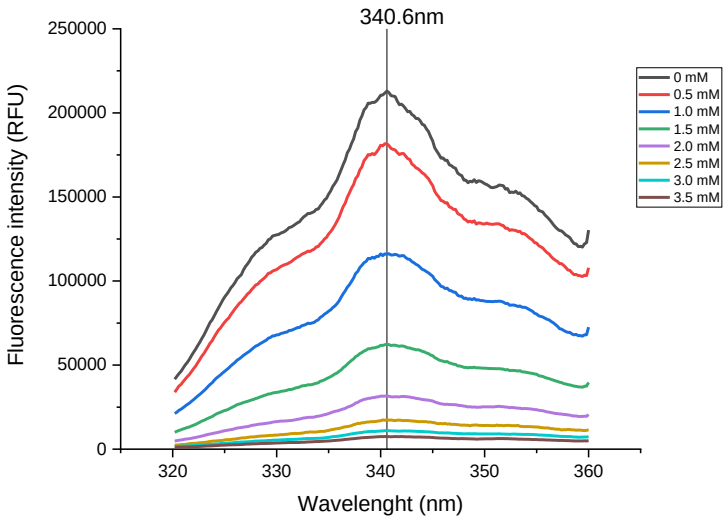


Fig. 3.20: Fluorescence emission spectra of tested protein in the absence of substrate (black line) and in presence of increasing concentrations of substrate 1 (0.5–3.5 mM) from black line onwards.

With that knowledge, the tryptophan fluorescence of both enzymes upon titration of substrates **1**, **2**, **4**, **5** and **6** was analyzed. Binding of these substrates to the proteins decreases its intrinsic fluorescence, indicating conformational changes in the microenvironment of the tryptophan residues. Plotting of the fluorescence data versus substrate concentration (**Fig. 3.21**) and fitting of the resulting titration curves using the non-linear regression (chapter 2.2.4.3, Eq.4) yielded binding constants K_D for each enzyme-substrate combination (**Tab. 3.14**). Binding experiments were performed in the absence of GSH to prevent substrate conversion.

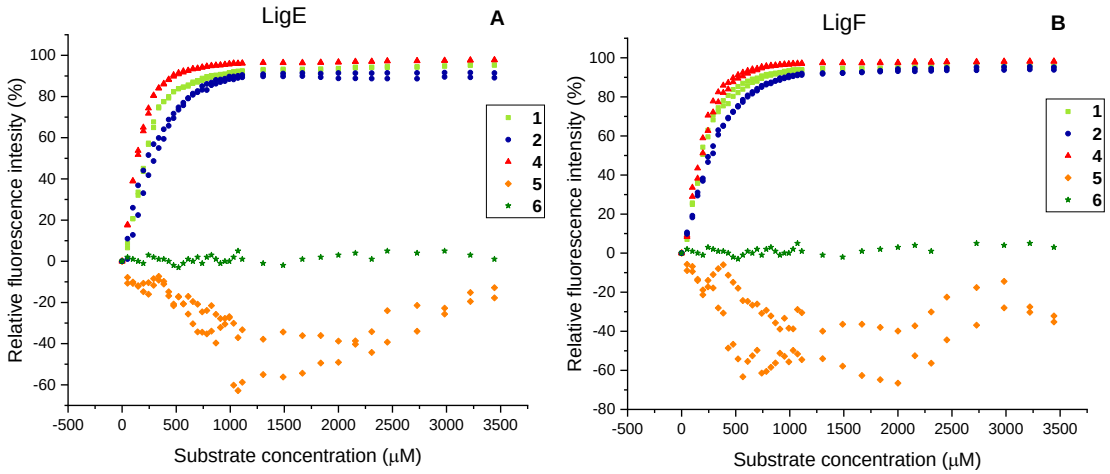


Fig. 3.21: Binding of lignin-model substrates **1**, **2**, **4**, **5**, and **6** to the active site of WT LigE (**A**) and WT LigF (**B**) as observed by tryptophan fluorescence. All measurements were performed in duplicate.

Tab. 3.14: Binding constants (K_D) and Hill coefficients (N) of LigE and LigF for binding of the model substrates **1**, **2**, **4**, **5**, and **6** resulting from tryptophan fluorescence measurements. All measurements were performed in duplicate.

Parameter	LigE					LigF				
	ligand					ligand				
	1	**2**	**4**	**5**	**6**	**1**	**2**	**4**	**5**	**6**
K_D (µM)	196±1.6	241±5.5	129±1.0	-	-	182±1.7	245±2.7	166±2.2	-	-
N	2.1	1.8	1.8	-	-	1.9	1.7	2.0	-	-

Based on the tryptophan fluorescence assay with the enzymes LigE and LigF, several interesting results could be observed. Substrate **6**, differing from model substrate **1** only in the presence of a hydroxyl group at the Cα carbon instead of a carbonyl group, yielded no changes in protein fluorescence during substrate titration. This suggests that this compound does not affect the surrounding of the tryptophan residues in the active site and thus, does not bind to the enzymes. This result was partially expected, since the tested β-etherases showed no catalytic activity towards substrate **6** (chapter 3.1.2) and generally, the Cα hydroxyl-containing substrates were proven not to undergo the desired bond cleavage [98–101,110]. A lack of affinity of the enzymes towards the Cα hydroxyl-

containing substrates could be one of the reasons for the absence of catalysis. In contrast, the titration of substrate **5**, in which the aryl ether bond of **1** was replaced by carbon-carbon bond, seemed to affect the intrinsic florescence of both enzymes, even though enzymes LigE and LigF showed no catalytic activity towards substrate **5** (chapter 3.1.2). The binding data received for the fluorescence binding assay however do not follow any obvious trend and therefore, it is difficult to conclude if binding of **5** to the active site occurs or not. Apart from that, clear binding of substrates **1**, **2** and **4** to both enzymes, LigE and LigF, was observed by tryptophan fluorescence. The data describing the titration of the standard model substrate **1** could be fitted to the Hill equation and thus quantify the binding affinity of the enzymes LigE and LigF towards the substrate. K_D values for LigE and LigF were found to be 196 and 182 µM, respectively. The ability of the enzymes LigE and LigF to convert substrate **1** (chapter 3.1.2) envisioned necessary binding of the substrate in catalytic site of these enzymes. The affinity of the enzymes LigE and LigF towards the achiral, side-chain-truncated substrate **2** yielded K_D values of 241 and 245 µM, respectively. Lower affinity of the enzymes towards the substrate **2**, when compared to substrate **1**, potentially explains lower catalytic activity observed in chapter 3.1.2. Surprisingly, data describing the titration of substrate **4**, where the aryl ether bond of **1** was replaced by thioether bond, revealed the K_D values for LigE and LigF to be 129 µM and 166 µM respectively. These results demonstrate stronger ligand-protein affinity compared to substrates **1** and **2**, even though catalysis with substrate **4** was not detected (3.1.2) due to the different properties of the thioether bond present in **4**. DFT calculations suggested that cleavage of the thioether bond in **4** would be an endothermic reaction, explaining the lack of enzyme activity, whereas cleavage of the corresponding ether bond is an exothermic process.

In all cases, where substrate binding was observed, a Hill coefficient N close to 2 was obtained. Both, LigE and LigF, are homo-dimeric enzymes. Hence, a Hill coefficient closed to 2 indicates that binding of the substrate to one active site influences substrate binding at the second active site, resulting in cooperative substrate binding.

Additionally, several LigE and LigF mutants with reduced or increased activity or fully inactive were analyzed by tryptophan fluorescence assay to investigate their K_D values for substrate binding (**Tab. 3.15**).

Tab. 3.15: Binding constants (K_D) and Hill coefficients (N) of LigE and LigF mutants for binding of model substrates **1**, **2**, and **4** in comparison to their previously determined relative activities. All measurements were performed in duplicate.

Enzyme	Ligand								
	1			**2**			**4**		
	K_D (µM)	N	Relative activity (%)	K_D (µM)	N	Relative activity (%)	K_D (µM)	N	Relative activity (%)
LigE	196±1.6	2.1	100	241±5.5	1.8	100	129±1.0	1.8	n.d.
LigE-Y122A	145±2.1	1.6	1.0±0.01	135±4.0	1.3	0.9±0.00	77±1.2	2.1	n.d.
LigE-Y23F	184±1.3	1.9	0.1±0.00	225±4.6	1.6	0.2±0.00	128±0.7	1.8	n.d.
LigE-R138A	142±1.5	2.0	n.d.	112±2.5	1.3	n.d.	99±1.4	2.1	n.d.
LigF	182±1.7	1.9	100	245±2.7	1.7	100	166±2.2	2.0	n.d.
LigF-F7A	177±1.1	1.9	0.6±0.12	287±3.7	1.8	0.5±0.01	174±1.8	2.0	n.d.
LigF-S111A	179±4.3	1.5	0.1±0.00	248±3.9	1.7	0.1±0.00	153±4.4	1.6	n.d.
LigF-W115A	170±2.1	1.9	n.d.	276±4.9	1.7	n.d.	167±1.5	2.1	n.d.
LigF-I122A	118±2.3	1.8	43±1.2	281±3.1	1.9	33±2.9	117±1.9	1.7	n.d.
LigF-P142A	190±3.0	1.7	149±2.2	284±3.3	1.8	123±5.8	210±2.0	2.2	n.d.
LigF-K147A	139±2.4	2.0	0.3±0.03	275±2.9	1.8	0.4±0.00	118±1.3	1.8	n.d.
LigF-Q39A	199±1.5	2.0	n.d.	267±3.6	1.8	n.d.	180±1.4	2.3	n.d.

n.d.= not detected

Interestingly, the obtained results indicate that the binding of substrates **1**, **2**, and **4** seemed rather unaffected in the tested mutants. This suggests that the mutants, despite the observed effects on activity, are still able to bind the substrates in the active site. It was noted, however, that the affinity for substrate **2** is in most cases lower compared to the other substrates. Additionally, LigF-I122A seems to have a higher affinity towards substrates **1** and **4** than WT LigF. Moreover, LigE mutants Y122A and R138A seem to have a higher affinity for all three tested substrates compared to WT. As described before for LigE and LigF WT, the binding of substrates **1**, **2**, and **4** to the mutants resulted again in Hill coefficients N of approx. 2, suggesting cooperative substrate binding.

Based on the LigF structure with docked substrate **1**, it was hypothesized that replacing the small amino acid Ala11, located at the entrance of the substrate-binding tunnel, by a large aromatic phenylalanine may affect access of the substrate to the active site and therefore influence K_D (**Fig. 3.22**). To confirm this hypothesis, mutant LigF-A11F was generated and analyzed. The resulting data is summarized in **Tab. 3.16**.

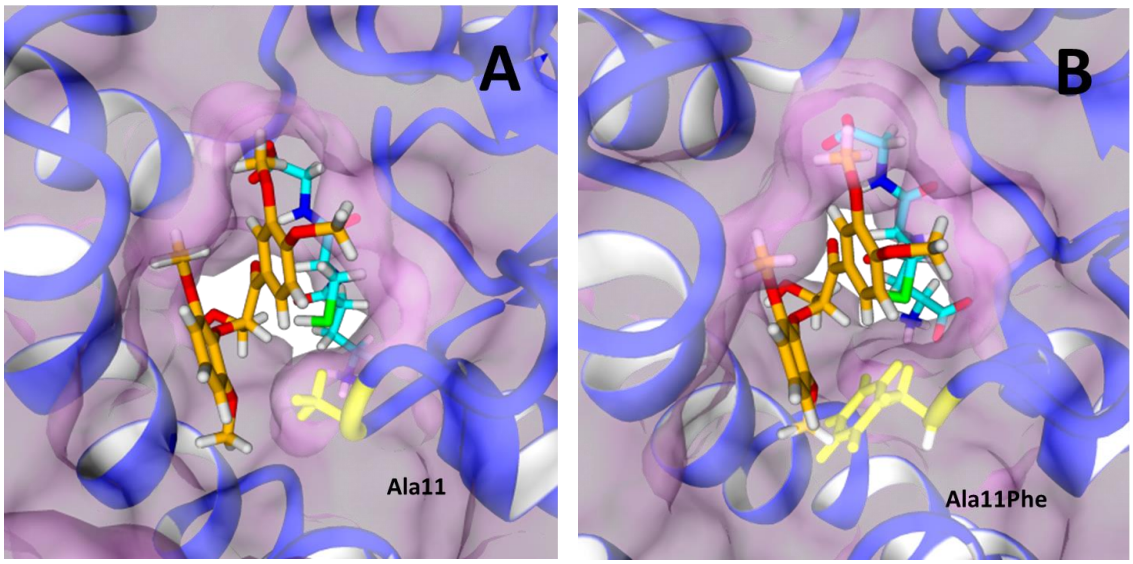


Fig. 3.22: Surface of the tunnel-like active site of LigF (Protein Data Bank code 4XT0) with model substrate **1** (orange) docked into the active site. GSH (cyan) and residue Ala11 (yellow), located at the possible entrance of the substrate-binding site are shown as well. (**A**) Wild-type LigF, (**B**) LigF mutant with Ala11 replaced by phenylalanine, possibly blocking the entrance of the substrate to the active site.

Tab. 3.16: Activity, kinetic parameters and K_D value of mutant LigF-A11F compared to WT LigF. Activity and kinetic measurements were performed in triplicate, tryptophan fluorescence measurements for substrate binding were performed in duplicate.

Enzyme	Relative activity towards		Kinetic parameters towards **3**				Thermodynamic parameters to **1**	
	1	**3**	K_M	K_I	k_{cat}	k_{cat}/K_M	K_D	N
	(%)		(mM)	(mM)	(s^{-1})	(mM^{-1} s^{-1})	(mM)	
LigF	100	100	0.93±0.16	3.6±0.97	4.8±0.56	5.3	0.18±1.69	1.9
LigF-A11F	0.5±0.01	2.0±0.01	9.4±3.59	0.53±0.22	1.3±0.42	0.1	0.18±1.88	1.8

The mutant A11F, compared to WT, displayed an obvious loss in catalytic activity (>99 % for **1** and >98 % for **3**) and catalytic efficiency (>97 %), confirming a clear negative effect of mutation A11F on substrate binding and conversion. On the other hand, substrate binding described by the thermodynamic parameter K_D indicated only a minor change (~10 %) when compared to wild-type LigF. These results suggest that, despite the ability of mutant A11F to bind the substrate in the active site, the positioning/arrangement of the substrate within the active site is altered and the conversion of the substrate is thus negatively affected. Furthermore, the K_I of the mutant A11F is 7-times lower than that of WT, causing a significant increase in substrate inhibition, which negatively affects the mutant's activity as well.

3.4.1.2 Substrate binding with GSH addition

All previous binding experiments with the tryptophan fluorescence assay were performed without cofactor GSH to prevent the reaction from occurring. We hypothesized, however, that substrate binding could be affected by the presence or absence of GSH in the

enzyme active site. To test this hypothesis, the tryptophan fluorescence assay was repeated in the presence of 1 mM GSH and 10 mM GSH (**Fig. 3.23**).

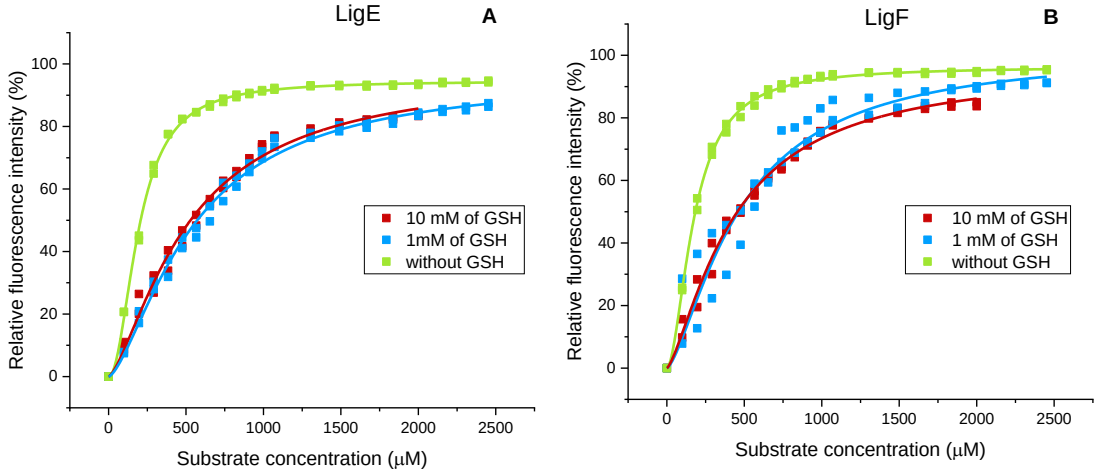


Fig. 3.23: Comparison of the binding curves of model substrate **1** to the active sites of LigE (**A**) and LigF (**B**) without GSH cofactor (green), with 10 mM GSH (red) and with 1 mM GSH (blue). All measurements were performed in duplicate.

Based on the obtained data, it is apparent that the binding affinity was negatively impacted in both enzymes in the presence of GSH. For LigE, the K_D value in reaction with 1 mM GSH and 10 mM of GSH is approx. 2.6 and 2.8-times higher, respectively, than in the reaction without GSH. Similarly, for enzyme LigF, the K_D value in the reaction with 1 mM GSH is approx. 2.4-times higher, while in the reaction with 10 mM of GSH the K_D is approx. 2.6-times higher than in the reaction without GSH. An evident decrease of substrate affinity was also observed in several LigE and LigF mutants as shown in **Tab. 3.17**.

Tab. 3.17: Dissociation constants K_D and Hill coefficients N of LigE and LigF wild type and mutants with model substrate **1** without GSH cofactor, with 10 mM GSH and with 1 mM GSH. All measurements were performed in duplicate.

Enzyme	Ligand **1**					
	without GSH		10 mM GSH		1 mM GSH	
	K_D (µM)	N	K_D (µM)	N	K_D (µM)	N
LigE	196±1.6	2.1	514±29.7	1.4	545±20.2	1.5
LigE-Y122A	145±2.1	1.6	371±19.2	1.4	419±12.4	1.6
LigE-Y23F	184±1.3	1.9	440±72.9	1.4	466±14.3	1.6
LigE-R138A	142±1.5	2.0	426±12.3	1.3	525±11.2	1.4
LigF	182±1.7	1.9	432±20.5	1.4	473±43.9	1.5
LigF-F7A	177±1.1	1.9	463±14.0	1.7	467±15.1	1.6
LigF-S111A	179±4.3	1.5	320±11.7	1.5	405±10.6	1.6
LigF-W115A	170±2.1	1.9	478±20.9	1.5	567±16.2	1.7
LigF-I122A	118±2.3	1.8	432±21.1	1.4	533±20.2	1.7
LigF-P142A	190±3.0	1.7	452±13.9	1.7	471±19.1	1.7
LigF-K147A	139±2.4	2.0	456±20.4	1.4	443±41.9	1.4
LigF-Q39A	199±1.5	2.0	437±16.8	1.6	520±14.1	1.6

A possible explanation of the effect of GSH on substrate binding is that the continuous conversion of the bound substrate in the active site with subsequent product release requires a higher substrate concentration in order to achieve a certain change in tryptophan fluorescence. A surprising observation was made for the mutants LigE-R138A and LigF-Q39A. Both mutated residues appear important for GSH binding in LigE and LigF, respectively. Despite the fact that both mutations caused a complete loss of enzymatic activity, likely caused by inefficient GSH binding, the addition of GSH during tryptophan fluorescence measurements still had a negative effect on binding of substrate **1**. Among all tested mutants and wild-type enzymes, no major difference in the effect of 1 mM and 10 mM GSH addition was observed, suggesting that 1 mM already caused full saturation of the binding pocket with GSH.

3.4.1.3 Binding of the cofactor GSH

A tryptophan fluorescence assay was also attempted to measure the dissociation constants of LigE and LigF for binding of the cofactor GSH, but no intrinsic fluorescence changes were observed when GSH was titrated to the proteins. This is likely explained by the lack of tryptophan residues in the GSH-binding pocket. However, it was speculated that the binding of GSH may be followed by tryptophan fluorescence when substrate is already bound in the enzyme active site. A tryptophan fluorescence assay was performed with addition of 0.2 mM and 2.0 mM substrate **1** (12.5 eq and 125 eq compared to the protein concentration), while GSH was subsequently titrated to the mixture. The results of the analysis shown in **Fig. 3.24** reveal a different behaviour for

LigE and LigF. While titration of GSH to LigE results in a clear trend in the change of fluorescence intensity, the fluorescence data of LigF upon GSH titration is only scattered.

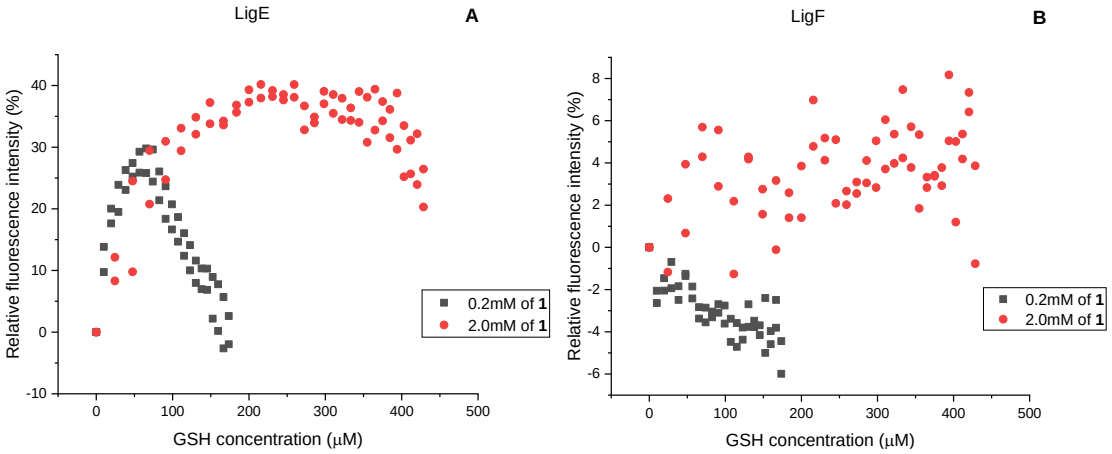


Fig. 3.24: (**A**) Tryptophan fluorescence of LigE upon titration of cofactor GSH in the presence of 0.2 mM (black) and 2.0 mM (red) model substrate **1**. (**B**) Tryptophan fluorescence of LigF upon titration of cofactor GSH in the presence of 0.2 mM (black) and 2.0 mM (red) model substrate **1**. All measurements were performed in duplicate.

The binding of GSH to LigE in the presence of 0.2 mM or 2.0 mM of **1** partially resembles a typical binding curve, though the observed fluorescence changes could not be fitted to equation **Eq.4**. A possible explanation for the curve shape observed with 0.2 mM **1** (black) is that at the beginning, with addition of 0-64 µM GSH, tryptophan fluorescence decreases (shown as steep increase of relative fluorescence intensity) due to GSH binding. This is followed by a fast increase in fluorescence (shown as decrease of the curve) at higher GSH concentration due to substrate conversion and depletion. Hence, the intrinsic tryptophan fluorescence goes back to the level of the enzyme without substrate present. In contrast, with 2 mM **1** (red) it takes more time and GSH to observe this fluorescence increase (described by a decrease of the curve), as more substrate is present. These results suggest that at low GSH concentration, the change in fluorescence intensity possibly describes the GSH binding behaviour, but the competing substrate conversion prevents from receiving a typical binding curve and hence, an accurate K_D value.

The fact that only scattered data is observed in the analysis with LigF could be explained by the structure of the enzyme. As shown in **Fig. 1.12**, the active site of LigF is located in a tunnel-like structure, which is partially closed from one side, as compared to the surface-exposed active site of LigE. Hence, prior substrate binding in the active site of LigF might prevent later GSH binding. Another possible explanation could be the lack of detectable fluorescence changes upon GSH binding to LigF.

3.4.2 Isothermal titration calorimetry (ITC)

The inability to quantify GSH binding to the β-etherases directly via tryptophan fluorescence assay prompted us to test a different method. ITC represents a direct, quantitative technique used to determine the thermodynamic parameters of interactions between small molecules (ligand) and larger molecules (protein). The complex formation between GSH and LigF protein was studied by this method. When GSH was titrated to free LigF, a dissociation constant of 80 µM was determined, which indicates a rather strong interaction (**Fig. 3.25A** and **Tab. 3.18**). The impact of 50, 100 and 200 µM of compound **4** on GSH binding in measurements with 100 µM of LigF protein was tested as well (**Fig. 3.25B** and **Tab. 3.18**). Compound **4** was chosen for this experiment due to its confirmed binding to the active site of LigF, while it cannot be converted by the enzyme. In the presence of lignin model substrate **4**, approx. 4-times higher K_D values were obtained for GSH binding. This is in agreement with the previously observed higher K_D values for lignin model substrate binding in the tryptophan fluorescence assay in the presence of GSH compared to measurements without GSH. Based on the received data, it appears that GSH binding is weaker in the presence of substrate compared to GSH alone. Similarly, binding of substrate seems to be weaker in the presence of GSH compared to substrate alone.

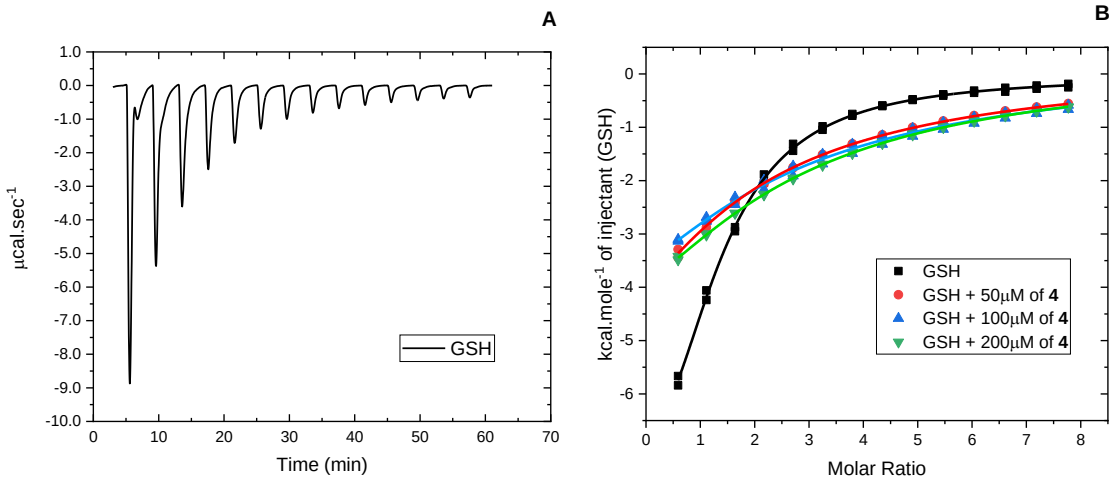


Fig. 3.25: (**A**) ITC measurement for the binding of GSH to wild-type LigF. The thermogram shows considerable heat release upon titration of the cofactor GSH to the LigF protein. (**B**) The ITC heat data, corrected for dilution, are plotted against the molar ratio of GSH to protein LigF with different amounts of substrate **4** present, and GSH. Sigmoidal curves typical of exothermic-binding reactions were obtained in all cases, however, changes in binding behaviour were observed in experiments with added substrate **4**.

Tab. 3.18: Dissociation constant K_D and coefficient N (referring to the number of binding sites) obtained from ITC measurements for the binding of cofactor GSH to wild-type LigF in the presence and absence of substrate **4**. All measurements were performed in duplicate.

Enzyme	GSH binding							
	no addition of **4**		50 µM of **4**		100 µM of **4**		200 µM of **4**	
	K_D (µM)	N	K_D (µM)	N	K_D (µM)	N	K_D (µM)	N
LigF	80±6.0	1.1	386±31.6	1.0	386±56.1	1.9	323±45.6	2.0

This observed change in affinity might be the result of structural changes within the GSH binding site upon substrate binding. At the same time, it was observed that the parameter N changed from approx. 1 in the measurements of GSH binding without analog **4** and in the presence of 50 µM **4**, to a value of approx. 2 in the presence of 100 µM and 200 µM of **4** (in the latter cases the substrate concentration was equal to or higher than the enzyme concentration). The parameter N in ITC analyses describes the stoichiometry of protein and ligand binding. Hence, at first glance, the increased N-value could suggest the presence of two GSH binding sites in LigF.

3.5 Non-lignin-based substrates

Previous characterizations of β-etherases and glutathione lyases were demonstrated on either lignin-like model substrates [104,107,108,110,112,131] or using various purities of polymeric lignin [27,119,120,122]. In contrast, the ability of these enzymes to cleave ether bonds in non-lignin like molecules could potentially increase their applicability. To test this, several compounds representing oligomers of manmade plastic material were provided by Prof. Dr. D. E. Kaufmann from the Institute of Organic Chemistry at TU Clausthal, Germany. Substrates **30**, **31**, **32**, **33**, and **34** were tested in reactions with β-etherases and glutathione lyases and are shown in **Fig**. **3.26**.

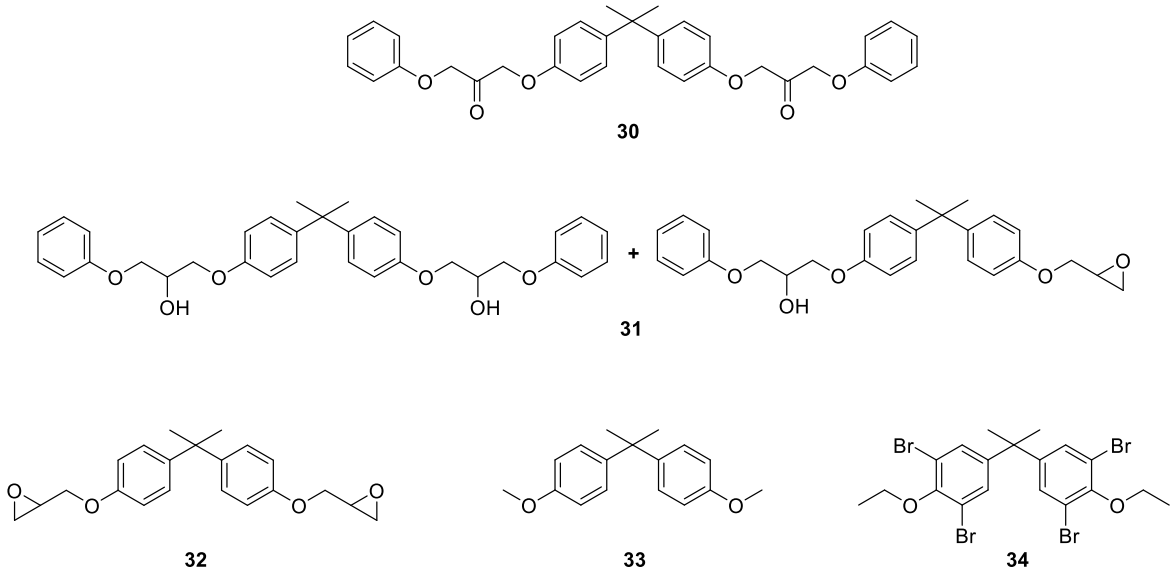


Fig. 3.26: Non-lignin-based substrates tested in reaction with β-etherases and glutathione lyases used in this work. [(**30**, 3,3'-((propane-2,2-diylbis(4,1-phenylene))bis(oxy))bis(1-phenoxypropan-2-one)), (**31**, mixture of 3,3'-((propane-2,2-diylbis(4,1-phenylene))bis(oxy))bis(1-phenoxypropan-2-ol) and 1-(4-(2-(4-(oxiran-2-ylmethoxy)-phenyl)propan-2-yl)phenoxy)-3-phenoxypropan-2-ol), (**32**, 2,2'-(((propane-2,2-diylbis(4,1-phenylene))-bis(oxy))-bis(methylene))bis(oxirane)), (**33**, 4,4'-(propane-2,2-diyl)bis(methoxybenzene)), (**34**, 5,5'-(propane-2,2-diyl)bis(1,3-dibromo-2-ethoxybenzene))].

3.5.1 Compound 30

Compound **30** (eluting at 18.13 min on the HPLC chromatogram) was initially assumed to be pure. However, the HPLC chromatogram of **30**, dissolved in pure DMSO, revealed a contamination eluting at 16.86 min (**Fig. 3.27**).

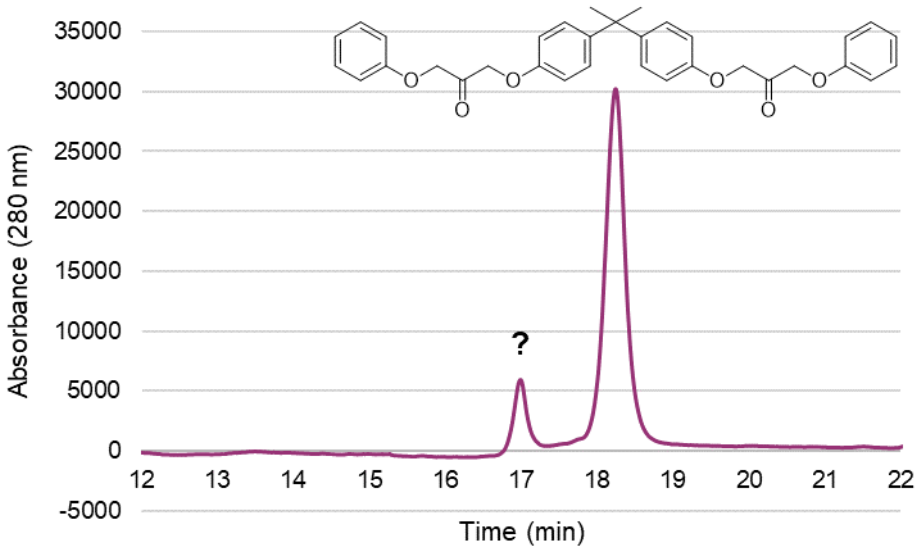


Fig. 3.27: Chromatogram of compound **30** dissolved in DMSO. Two peaks are observed, a minor peak at 16.86 min and a major peak at 18.13 min, of which the structure is depicted in the figure.

Based on the structure of the major compound of mixture **30**, it was hypothesized that the ether bonds present could possibly be cleaved by β-etherases. Reactions were

performed at 25°C using all available β-etherases (LigE, LigE-NS, LigE-NA, LigP, LigF, LigF-NS, LigF-NA) and consisted of 0.2 mM of non-lignin substrate, 1 mM GSH and 50 µg of tested enzyme. In order to efficiently dissolve the substrate and still ensure catalytic activity of the enzymes, 10 % v/v of DMSO was used as co-solvent in the reaction (compared to standard 5 % v/v of DMSO used with lignin model substrates). Despite the increased solvent concentration, substrate **30** as well as the other tested non-lignin-like substrates were not completely dissolved under these conditions, complicating sample drawing and subsequent HPLC analysis. Hence, an alternative approach for sample drawing and analysis was followed. Instead of preparing a reaction of bigger volume to take several samples over time, 10 parallel reactions with the same composition (along with respective control reactions) have been prepared and started simultaneously. For sample drawing at different time points, always two parallel reactions were stopped at the same time by dilution with DMSO in 1:10 ratio. This also ensured complete dissolution of the substrates and possible products, and consecutive centrifugation separated precipitated protein. Received samples were analyzed by HPLC. With this approach it was possible to follow the reaction progress over time (five individual time points). Example chromatograms of a respective reaction of substrate **30** and the non-enzymatic control reaction are shown in **Fig. 3.28**.

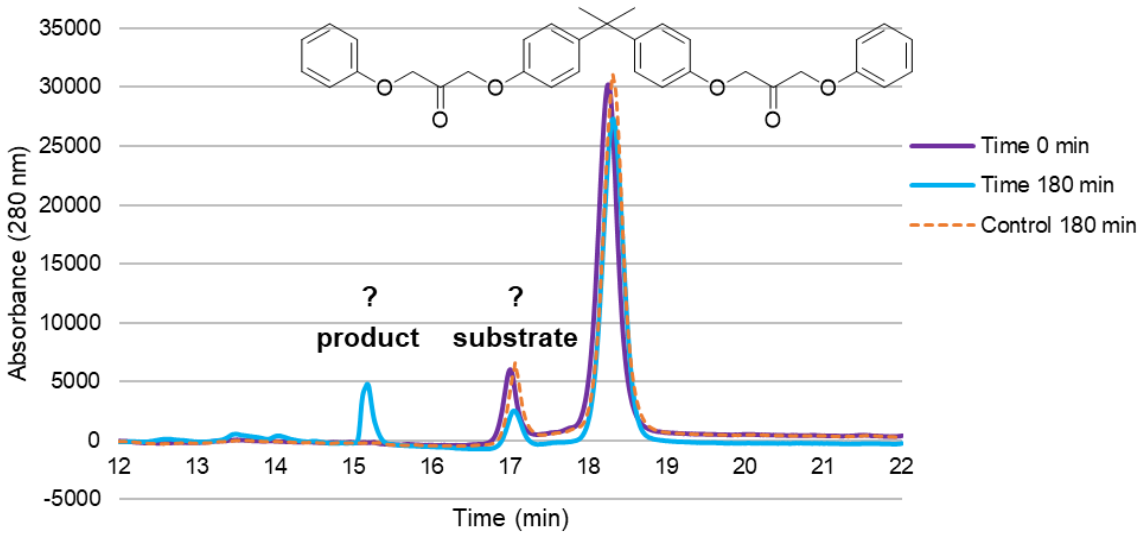


Fig. 3.28: Comparison of the chromatograms at time point 0 min (purple), after 180 min (blue) and the non-enzymatic control reaction (dashed orange) of a reaction with substrate **30** and LigE-NA as catalyst. A visible decrease of the peak at 16.86 min was accompanied with the appearance of a new peak at 15.09 min. The major peak at 18.13 min remained relatively unchanged. No activity was observed in control reaction.

Fig. 3.28 depicts the reaction results using β-etherase LigE-NA, but enzymes LigE, LigF, LigF-NS and LigF-NA showed identical peak formation (data not shown) with various

levels of activity. Enzymes LigE-NS and LigP showed no activity and no background reaction was observed in control reaction without enzyme. Based on the peak changes, it was assumed that the peak eluting at 16.85 min represented the substrate of the reaction, while the new peak at 15.09 min represented the formed product. Since the substrate peak (16.85 min) never fully disappeared, it was hypothesized that the substrate might be racemic. However, the combination of the two most active enantio-complementary β-etherases (LigE and LigF-NA) in the reaction did not result in complete conversion of the substrate (data not shown), and thus this hypothesis was not confirmed. In order to identify the substrate and the product, MS analysis was performed in TU Braunschweig facilities with help from Dr. Uli Papke. The obtained MS results (shown in chapter 7.6) enabled identification of the potential substrate and product of the reaction as shown in **Fig. 3.29**.

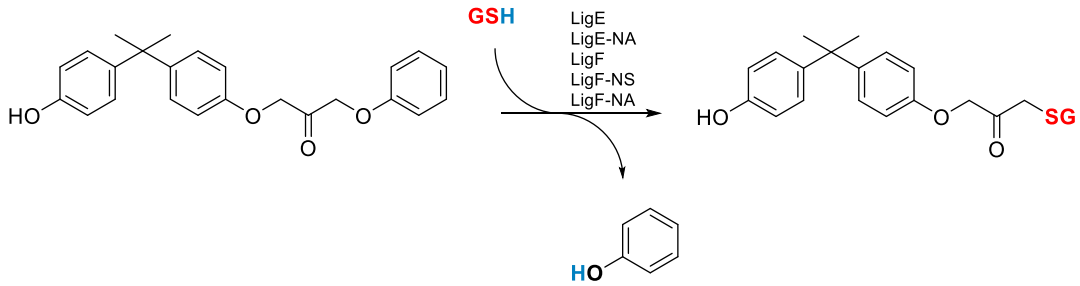


Fig. 3.29: Scheme of the reaction catalyzed by β-etherases based on mixture **30** as proposed by the MS results.

The ether bond of the substrate was successfully cleaved with release of phenol and the glutathione adduct. In order to investigate a further possible conversion of the formed glutathione adduct, glutathione lyases LigG, LigG-NS and LigG-TD together with fresh GSH were added to the reaction, however, no further cleavage occurred (data not shown). Despite the fact that a potential novel substrate for cleavage of the ether bond by β-etherases was identified, further analysis will be necessary. NMR data of substrate and product could not be obtained due to the limited amount of received mixture **30**.

3.5.2 Mixture 31

Mixture **31** was expected to contain two major compounds (eluting at 19.20 and 20.28 min on the HPLC chromatogram in **Fig. 3.30**) with known structures. Both structures represent intermediates in the chemical synthesis of compound **30**. However, the HPLC chromatogram of mixture **31**, dissolved in pure DMSO, showed again an unknown contamination eluting at 18.75 min.

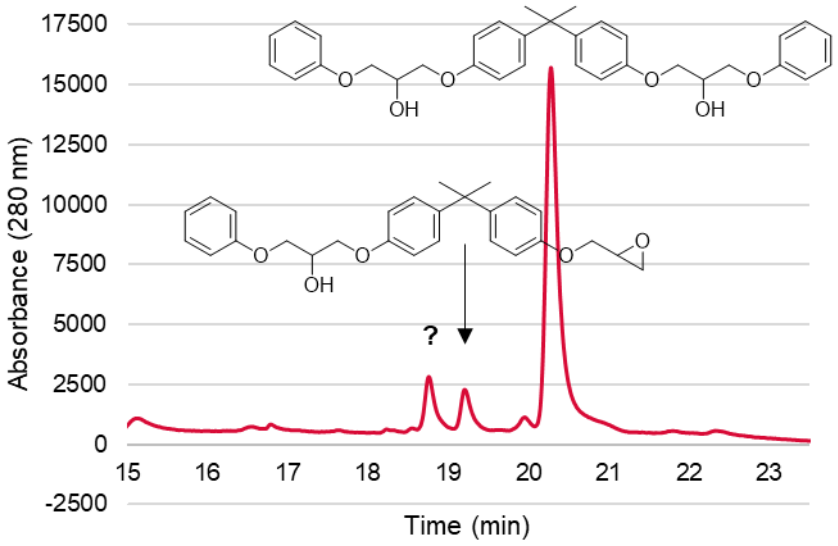


Fig. 3.30: HPLC chromatogram of mixture **31** dissolved in DMSO. Three peaks are observed, an unknown one at 18.75 min and two expected ones at 19.20 min and 20.28 min, for which the structures are depicted in the figure.

The major compound of mixture **31** (with retention time 20.28 min) is structurally similar to compound **30**, except for the presence of hydroxy groups at the β-carbons (relative to the ether bonds) compared to keto groups in compound **30**. The other known compound of mixture **31** (with retention time 19.20 min) is structurally similar but additionally carries an epoxide ring. Reactions with mixture **31** as a possible substrate were performed using all available β-etherases (LigE, LigE-NS, LigE-NA, LigP, LigF, LigF-NS, LigF-NA) identically to the pervious reactions with substrate **30**. A representative HPLC chromatogram after enzymatic catalysis together with the non-enzymatical reaction are shown in **Fig. 3.31**.

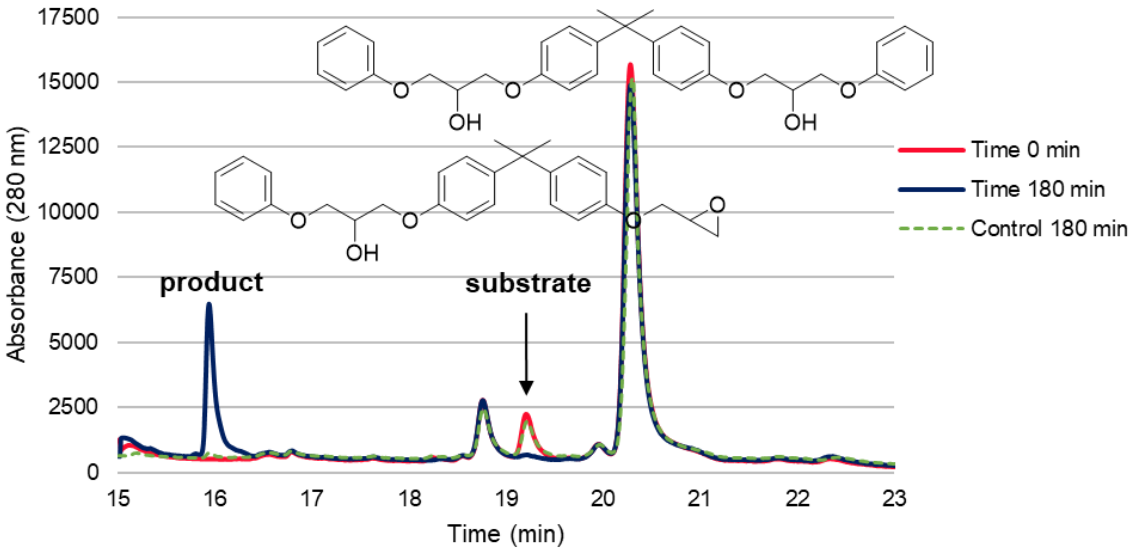


Fig. 3.31: Comparison of the chromatograms at time point 0 min (red), after 180 min (plum) and the non-enzymatic control reaction (dashed green) of a reaction with mixture **31** as substrate and LigE-NA as catalyst. A visible decrease of the peak at 19.20 min was accompanied with the appearance of a peak at 15.94 min. The peaks at 18.75 min and 20.28 min remained unchanged. No activity was observed in control reaction.

Based on the chromatogram shown in **Fig. 3.31**, it was assumed that during the reaction the compound represented by the peak at 19.20 min was converted by the β-etherases with formation of a new product eluting at 15.94 min. The other two peaks at 18.75 min and 20.28 min remained unchanged during the reaction. Similar results were obtained for reactions catalyzed by LigE, LigE-NS, LigE-NA, LigF, LigF-NS and LigF-NA. In contrast, the enzyme LigP showed no activity and no background activity was observed in control reaction without enzyme. This time, the substrate structure was known but the product structure had to be determined. In order to identify the substrate and the product, LC-MS analysis was performed in collaboration with the group of Prof. Dr. Mark Brönstrup (Department of Chemical Biology at the Helmholtz Centre for Infection Research (HZI) Braunschweig). The obtained LC-MS results (shown in chapter 7.6) enabled identification of the putative reaction product as shown in **Fig. 3.32**.

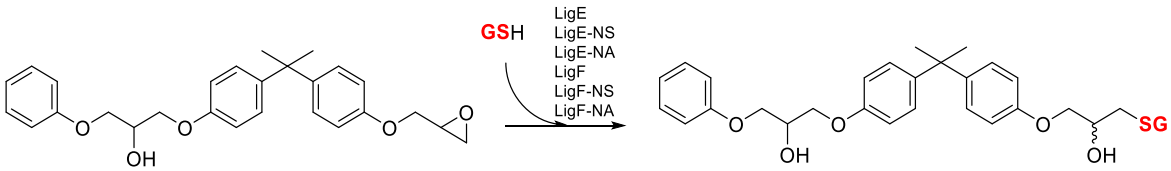


Fig. 3.32: Scheme of the reaction catalyzed by β-etherases starting from mixture **31** as proposed by the LC-MS results.

This time, no ether bond was cleaved but the epoxide ring was opened by nucleophilic attack of glutathione. Such a reaction has not been described before for β-etherases,

although a variety of mammalian glutathione S-transferases, including enzymes from the alpha, pi and mu class, have been reported to catalyze epoxide ring opening on a variety of substrates forming corresponding GSH-conjugates [149–151]. Because of the limited amount of the substrate, the reaction could not be explored further and no NMR data of substrate and product could be obtained.

Subsequent conversion of the formed glutathione conjugate with glutathione lyases (LigG, LigG-NS, and LigG-TD) was tested as well, but no further conversion was observed.

3.5.3 Substrates 32, 33 and 34

Substrates **32**-**34** were received as pure compounds. Since all three contained either ether bonds or epoxide rings, it was hypothesized that they could represent new substrates for β-etherases. The substrates were tested using all available β-etherases (LigE, LigE-NS, LigE-NA, LigP, LigF, LigF-NS, LigF-NA). Although even increased amounts of enzymes were used, no conversion of substrates **33** and **34** could be observed (data not shown). The only substrate which was converted during the reaction was compound **32**, for which the respective HPLC chromatogram is shown in **Fig. 3.33**.

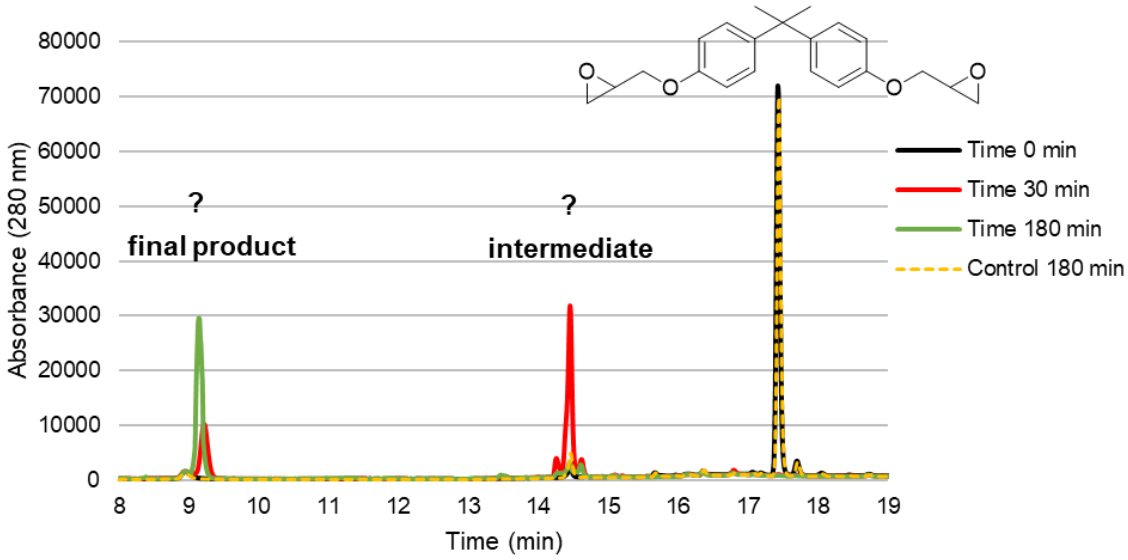


Fig. 3.33: Comparison of the chromatograms at time points 0 min (black), 30 min (red), 180 min (green) and non-enzymatic control reaction (dashed yellow) of a reaction with substrate **32** and LigE-NA as catalyst. A visible disappearance of the substrate peak at 17.43 min was observed within 30 min of the reaction. Formation of a new peak at 14.15 min was visible at time point 30 min, which decreased again with further reaction. After 180 min of the reaction the peak had disappeared, suggesting it to be an intermediate of the reaction. At the same time, a second new peak appeared at 9.17 min which further increased until time point 180 min and seems to represent the final product. In control reaction, slight formation of peak at 14.15 min was observed.

Based on the chromatograms of the reaction, substrate **32** (eluting at 17.43 min) is immediately converted to an unknown intermediate (eluting at 14.45 min), which is further converted to the final product (eluting at 9.17 min) of the reaction. Identical results (with various levels of activity) were obtained using β-etherases LigE, LigE-NA, LigF, LigF-NS and LigF-NA. In reactions with enzymes LigP and LigE-NS no activity towards substrate **32** could be detected. The non-enzymatic control reaction shows minor peak eluting at 14.15 min assigned to the background reaction possibly caused by high glutathione concentration. No background reaction was observed at the peak eluting at 9.17 min. The resulting product was further tested for possible conversion by glutathione lyases, although no further conversion was observed. LC-MS measurements were performed to identify the formed intermediate and the final product of the β-etherase-catalyzed reaction. The obtained LC-MS results (shown in chapter 7.6) enabled identification of both compound structures as shown in **Fig. 3.34**.

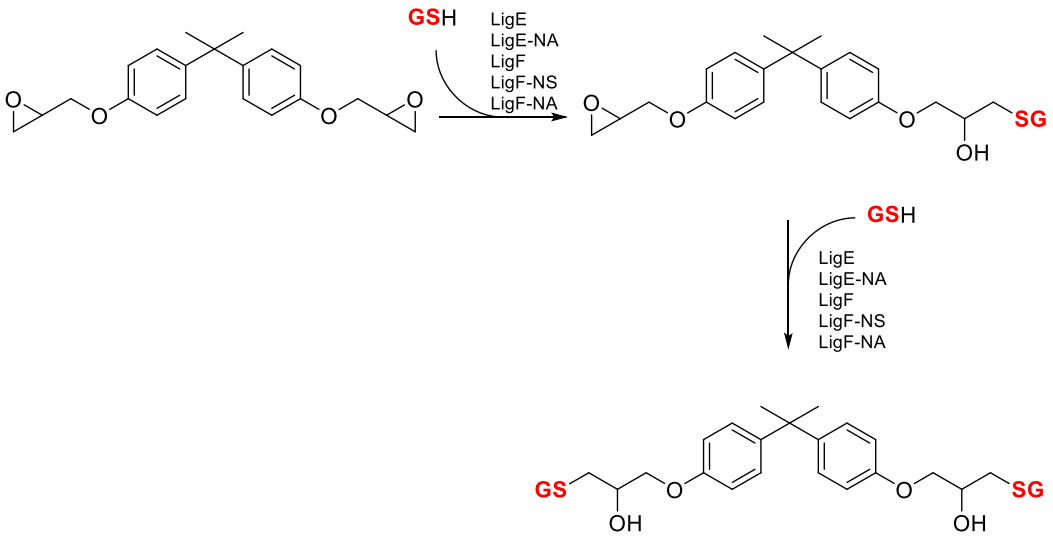
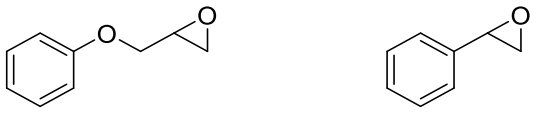


Fig. 3.34: Scheme of the reaction catalyzed by β-etherases starting from mixture **32** as proposed by the LC-MS results.

Since substrate **32** contains two epoxide rings, both can be attacked by glutathione. Accordingly, in the intermediate of the reaction only one epoxide ring is opened by nucleophilic attack of glutathione, whereas the final product has both epoxide rings opened with glutathione. Based on these results, it can be assumed that β-etherases are able to open epoxide rings of arylaliphatic compounds. Although the LC-MS data are convincing, further NMR analyses of the intermediate and the final product would be necessary.

Based on the previous results, two further arylaliphatic epoxides available in the laboratory were tested as potential substrates for β-etherases (**Fig. 3.35**).



7 **8**

Fig. 3.35: Schemes of arylaliphatic epoxide substrates tested for conversion by β-etherases [(**7**, 2-(phenoxymethyl)-oxirane), (**8**, 2-phenyloxirane)].

Unfortunately, epoxides **7** and **8** were not converted by any of the tested β-etherases. The control reactions with no enzyme added, however, already shown a strong background reaction (data not shown). This was rather unexpected as in previous reactions with epoxide substrates no or only a limited chemical background reaction was observed. In case of reactions with substrates **7** and **8**, a relatively high concentration of GSH in the aqueous environment probably caused the high chemical background of GSH-mediated epoxide ring opening, as described by Brunmark and Cadenas [152].

4 Discussion

In this work, the potential of different glutathione-dependent enzymes was explored in various settings. The investigated β-etherases and glutathione lyases have a huge potential for the cleavage of ether bonds in lignin in a predictable manner and thus, for the production of aromatics from lignin. The strict stereoselectivity of β-etherases makes these enzymes attractive for various applications in which enantiopure compounds are envisaged. To overcome the economically unattractive cofactor requirement, a whole-cell biocatalyst with glutathione regeneration was established, as described in the first part of this thesis. Co-expression of the most active β-etherases with complementary stereoselectivity (LigE and LigF-NA) and a glutathione lyase converting both enantiomers of the resulting glutathione adduct (LigG-TD) in *E. coli* provided a whole-cell catalyst with efficient GSH recycling.

In the second part of this thesis, the best-studied β-etherases LigE and LigF underwent further analysis, in which active site residues of both enzymes were subjected to mutagenesis to investigate their impact on substrate binding and enzyme activity. In both cases, most of the introduced mutations had a negative impact on catalytic activity, while some of the LigF mutants displayed increased activity.

In the last part, the ability of β-etherases to convert different manmade, non-lignin-derived compounds was investigated.

4.1 Heterologous expression and enzyme activity

To date, β-etherases and glutathione lyases have been described in various publications. In these, different *E. coli* strains were used for heterologous expression of these enzymes: *E. coli* B834 [107,109,115], *E. coli* MV1190 [97] or *E. coli* BL21 (DE3) [105,110,119,122]. In this thesis, the *E. coli* BL21 (DE3) strain was used for enzyme expression. Yields of purified β-etherases and glutathione lyases are comparable with those of Picart *et al*. [105,110], which is explained by use of the same expression strain and a similar expression protocol (LB broth as inoculum media in this thesis compared to TB media in Picart *et al*. [105,110], and expression at 20˚C for both, β-etherases and glutathione lyases compared to 25˚C for the glutathione lyases in Picart *et al*. [105]). In an experiment of Jan Terbrack [270] (bachelor student, 2018), however, the expression level of LigE and LigF could be increased by 56 % and 190 %, respectively, through addition of 2 mM $MgSO_4$ to the cultivation (expression) media [270]. Several publications discuss possibilities of how to improve the heterologous expression of proteins in *E. coli* [153–157]. These suggest that the concentration of magnesium ions in typical expression media is a limiting factor for

the sugar metabolism of the cells and thus affects culture growth and enzyme expression. The results obtained by Jan Terbrack in his bachelor thesis [270] support this and demonstrate that even higher yields of β-etherases and glutathione lyases are possible if the expression medium is further optimized.

Moreover, the produced β-etherases and glutathione lyases have been biocatalytically characterized towards different model dimeric substrates. The specific activity values obtained in this thesis differ significantly from previously published data [105,110] (**Tab. 4.1**). Other enzyme properties, such as stereoselectivity (LigE, LigE-NS, LigE-NA and LigP are strictly (*R*)-selective while LigF, LigF-NS and LigF-NS are strictly (*S*)-selective; LigG, LigG-NS and LigG-TD all possess (*R*)-stereopreference), are in agreement with the published data.

Tab. 4.1: Comparison of measured activity values for β-etherases and glutathione lyases with literature values by Picart *et al*. [105,110].

Enzyme	Substrate	Published specific activity (U/mg)	Measured specific activity (U/mg)	Difference (%)
LigE	*rac*-**1**	5.93 [110]	1.75	-71
LigE-NS	*rac*-**1**	2.78 [110]	1.00	-64
LigE-NA	*rac*-**1**	2.67 [110]	0.73	-73
LigP	*rac*-**1**	0.10 [110]	0.05	-50
LigF	*rac*-**1**	2.60 [110]	1.26	-52
LigF-NS	*rac*-**1**	0.37 [110]	1.60	+332
LigF-NA	*rac*-**1**	6.79 [110]	2.40	-65
LigG	(*R*)-**11**	1.74 [105]	5.40	+210
LigG	(*S*)-**11**	0.09 [105]	0.60	+567
LigG-NS	(*R*)-**11**	71.4 [105]	16.9	-76
LigG-NS	(*S*)-**11**	0.02 [105]	0.40	+1900
LigG-TD	(*R*)-**11**	36.2 [105]	13.3	-63
LigG-TD	(*S*)-**11**	0.12 [105]	6.20	+5067

There are several reasons that could explain the large disparity between the herein measured activity data and the published values. A major difference in the activity measurement was the applied enzyme concentration. While a constant amount of 10 µg of the β-etherases was used by Picart *et al*.[110], lower amounts of β-etherases (0.5-2 µg) were commonly used in our case to achieve a more accurate and linear product formation as the basis for specific activity determination. Moreover, the activity measurements for β-etherases and glutathione lyases were performed in a stronger buffer (100 mM) compared to Picart *et al*. (20 mM for β-etherases [110] and 50 mM for glutathione lyases [105]). This may have had a direct impact on enzyme activity but might have also ensured a more constant pH during the reaction. Despite the differences in absolute activity values, the order of β-etherases with increasing activity is still the same, describing LigE as the most active (*R*)-selective β-etherase and LigF-NA as the most

active (*S*)-selective β-etherase. Likewise, glutathione lyase LigG-TD is still the enzyme with the lowest stereopreference.

4.2 Whole-cell biocatalyst with GSH recycling

In order to implement β-etherases and glutathione lyases on an industrial scale, not only increased protein yields are necessary. One of the major obstacles for these enzymes is their dependence on the relatively expensive cofactor glutathione (as of March 2020, the price of reduced glutathione varied from 200 USD/kg [271] to 5190 USD/kg [272] depending on the purity and the country of origin). In favour of avoiding this requirement, Reiter *et al*. [119] used an NADH-dependent glutathione reductase from *Allochromatium vinosum* in combination with enzymes involved in lignin degradation (NAD^+-dependent Cα dehydrogenase LigD as well as β-etherase LigF and glutathione lyase LigG) to achieve GSH recycling. This way, Reiter *et al*. [119] were able to setup a self-sufficient biocascade for the cleavage of lignin model substrates with internal regeneration of both cofactors, NAD^+ and glutathione. A different approach was followed in this thesis to achieve GSH recycling intracellularly by *E. coli*-innate, NADPH-dependent glutathione reductase. Based on our research, as of February 2016 (beginning of this project), the most active enantiocomplementary β-etherases were LigE ((*R*)-selective) and LigF-NA ((*S*)-selective), and the glutathione lyase with the lowest stereopreference (to convert both enantiomers of the formed glutathione adduct) was LigG-TD (*R*/*S* = 2). Hence, these enzymes were chosen to be co-expressed within a single *E. coli* cell.

E. coli cells contain a significant amount of NADPH-dependent glutathione reductase that catalyzes the reduction of oxidized glutathione (GSSG) into two molecules of GSH. GSH and NADPH are produced by the *E. coli* metabolism and are important for maintaining a reducing environment within the cell [158]. Hence, the *E. coli* cell metabolism was utilized to provide and regenerate GSH and NADPH required for our enzyme cascade consisting of LigE, LigF-NA and LigG-TD (**Fig. 4.1**).

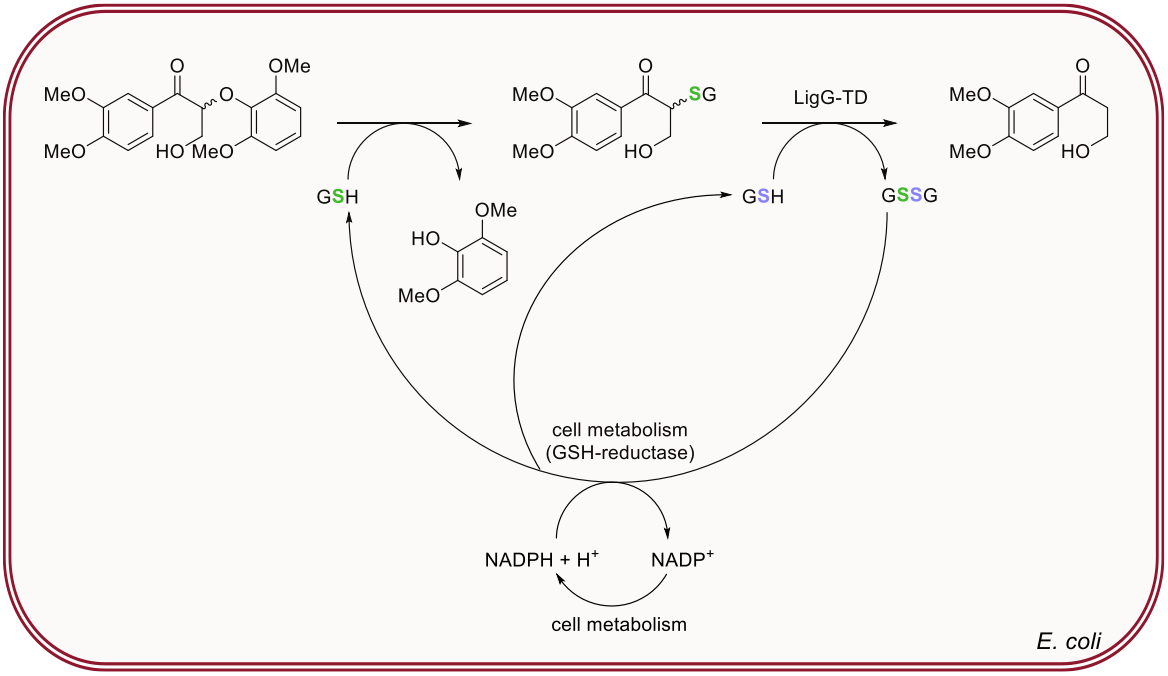


Fig. 4.1: Biocascade of GSH-dependent enzymes for ether bond cleavage, co-expressed in *E. coli* cells to ensure efficient GSH provision and recycling by the cell metabolim.

This whole-cell approach was tested with two *E. coli* strains: BL21 (DE3) [159] and C43 (DE3) [160,161]. *E. coli* BL21 (DE3) is commonly used in laboratories for recombinant protein production. The typical lack of lon and ompT proteases together with harboring the prophage DE3, derived from a bacteriophage λ, makes it probably the most widely used strain in high-level T7 expression of recombinant proteins [162,163]. The C43 (DE3) strain was derived as a double mutant from the parental BL21 (DE3) strain. The C43 (DE3) strain is proven to grow to high saturation cell density and is also able to over-express toxic proteins. Both strains were selected because they are highly suitable for expression of our GSH-dependent β-etherases and glutathione lyases.

The activity of both generated whole-cell biocatalysts was tested in the conversion of substrate **1**. As a result, full conversion of substrate **1** was only achieved with the *E. coli* C43 (DE3) (pETDuet1_*ligE*_ *ligF*-NA) (pIT2_*ligG*-TD) strain after 5 h of reaction compared to the *E. coli* BL21 (DE3) (pETDuet1_*ligE*_ *ligF*-NA) (pIT2_*ligG*-TD) strain, in which a conversion of 52 % was observed after 20 h of reaction. This could be explained by the difference in expression of proteins in each strain. While in the BL21(DE3) strain, the T7 RNA polymerase gene is under the control of the powerful, IPTG-inducible *lac*UV5 promoter, the strain C43(DE) contains three mutations in the *lac*UV5 promoter, turning this promoter into the much weaker wild-type lac promoter, ensuring a reduced transcription rate in C43 (DE3) [164,165]. As a result, proteins in BL21 (DE3) are expressed at a faster rate, making them more prone to aggregation and inclusion bodies formation

compared to the slower expression in the C43 (DE3) strain, securing proper enzyme folding [166,167]. Additionally, the C43 (DE3) strain is known for improved plasmid stability, which could lead to improved expression of the proteins as well [160]. Unfortunately, expression levels of the soluble β-etherases and glutathione lyase by BL21 (DE3) and C43 (DE3) strains have not been compared.

The constant interest of the research community in the field and the potential application of these enzymes led to multiple discoveries by other researchers within the timeframe of the experiment. In a publication about database mining for novel bacterial β-etherases [111], enzymes displaying even higher activity than LigE and LigF-NA have been identified. The LigE-type β-etherase with (*R*)-stereoselectivity from *Altererythrobacter* sp. 66-12 (GeneBank number: OJU60283) displayed ~ 2.9-times higher activity than LigE from *Sphingobium* sp. SYK-6, while the LigF-type β-etherase with (*S*)-stereoselectivity from *Altererythrobacter* sp. Root672 (GeneBank number: WP_055919008) exhibited ~2.3-times higher activity than LigF-NA from *N. aromaticivorans*. Similarly, progress was also made on the available range of glutathione lyases. The analysis by Kontur *et al*. [109] of the phylogenetic three of GSTs, known or predicted to catalyze reactions involved in the sphingomonad pathway for breaking the β-aryl ether bond, described various novel glutathione lyases with low stereopreference. One of them, $NaGST_{Nu}$, from *N. aromaticivorans* (GeneBank number: WP_011446237), was tested by Till Peters [273] (bachelor student, 2019) within his bachelor project and compared to LigG-TD (**Tab. 4.2**).

Tab. 4.2: Comparison of measured activity values and stereopreference for glutathione lyases LigG-TD (used in the whole-cell biocatalyst) and $NaGST_{Nu}$.

Enzyme	Specific activity towards *rac*-**11** **(**U/mg)	Specific activity towards (*R*)-**11** (U/mg)	Specific activity towards (*S*)-**11** (U/mg)	Rate (*R*) / Rate (*S*)
LigG-TD	11	13.3	6.2	2.1
$NaGST_{Nu}$	184	190	152	1.3

This new $NaGST_{Nu}$ not only exhibits an improved *R*/*S* ratio, suggesting that this glutathione lyase is the least stereospecific so far, but also the specific activity towards *rac*-**11** is almost 17-times higher. Hence, more active and less stereoselective enzymes could be applied in our whole-cell approach for internal GSH recycling in the future.

4.2.1 Optimal conditions for production and application of the whole-cell catalyst

Different conditions for protein expression were tested to yield a whole-cell catalyst with the highest activity in the conversion of **1** towards the final product VG. From the tested

expression temperatures (17˚C-37˚C), 20˚C resulted in the highest activity, suggesting that a lower temperature negatively impacted the level of protein expression. In contrast, higher expression temperatures should lead to faster expression of the recombinant proteins, which can be accompanied by protein aggregation via hydrophobic patches and hence formation of inclusion bodies if the newly formed proteins do not fold fast enough.

Optimal conditions for the whole-cell catalyzed reaction were also identified. The optimal extracellular pH for the reaction was investigated. The pH optimum for conversion of **1** using the whole-cell catalyst was observed at pH 8, whereas all three heterologously expressed enzymes display pH optima in the more alkaline range (LigE and LigF-NA at pH 9 [111] and LigG-TD at pH 9.5 [105]). This difference is not surprising, as the applied extracellular pH will not directly influence the intracellular pH, since *E. coli* cells will keep a rather constant intracellular pH independent of the extracellular conditions. It is reported that the intracellular pH of *E. coli* at optimal conditions is 7.6 [168]. The whole-cell reaction pH optimum of 8 therefore represents an extracellular pH value in which the *E. coli* metabolism is able to ensure satisfactory regeneration of GSH and other essential processes, while allowing a nucleophilic attack of GSH on the β carbon of the **1** catalyzed by β-etherases, for which the alkaline environment is desired.

Using model substrates for the investigation of β-etherases and glutathione lyases has proven useful due to the simple analysis of both reaction steps of the cascade by HPLC and their relatively simple organic synthesis, while representing the common β-*O*-4 arylether bond of “real” heteropolymeric lignin [20]. One of the obstacles that are connected with this substrate, and also represents a major problem with polymeric lignin, is its low solubility, which requires addition of cosolvent. The choice of organic solvent plays an important role in biocatalysis; not only is cosolvent addition necessary to dissolve hydrophobic substrates in the aqueous reaction mixture, it was also shown that in some cases the selection of organic solvent may affect the enzyme’s chemoselectivity, stereoselectivity and regioselectivity [169,170]. Additionally, the added cosolvent facilitates substrate uptake through the outer and inner membrane into the cytosol of *E. coli*, where catalysis and GSH regeneration occur, as it makes the cell membrane more permeable. A negative impact of increased solvent concentration on the whole-cell performance was confirmed for all of the tested solvents. Isopropyl alcohol as a solvent in low concentrations (0.5 % and 2.5 % v/v) appeared to have the lowest negative impact on the relative activity of the whole-cell catalyst. This was rather unexpected based on published data [133] reporting that aliphatic alcohols have the most damaging effect on the *E. coli* membrane. At higher solvent concentration (≥2.5 % v/v), the negative effect of

isopropyl alcohol was more apparent since the relative activity of the catalyst decreased. This negative impact of isopropyl alcohol is caused by its hydrophobic character, resulting in a slow permeabilization of the cell membrane with a detrimental effect on the cell metabolism [137] which is necessary for NADPH and GSH regeneration. In contrast, methanol as a solvent yielded the highest conversions with larger amounts of cosolvent in the reaction (≥2.5 % v/v). Overall, methanol had the lowest negative impact on the relative activity of the whole-cell catalyst. It was previously reported that methanol as a short-chain alcohol has only a weak ability to interfere with the lipid bilayer of a cell and thus is only minimally affecting membrane structure [171,172]. Methanol has been proven to have only low toxicity on *E. coli* growth compared to different other alcohols [133,138]. Up till now, several *E. coli* strains were engineered not only to tolerate high concentrations of methanol during growth and in the resting phase, but also to use methanol as a carbon source [173–177].

DMSO, as commonly used cosolvent and even cryoprotectant for long-term storage of bacteria was expected to yield the highest relative activity of the whole-cell catalyst. To our surprise, however, the performance of the whole-cell reaction in combination with DMSO was rather mediocre. Lower activities of the catalyst in reactions where DMSO was used as solvent might be explained by its stiffening effect on the fluidity of the water-liquid interface of the cell membrane as well as the non-polar part of the membrane [178]. Stiffening the phospholipid layer of the cells may hinder the diffusion of substrate through the cell membrane. On contrary, methanol was shown to slightly increase the cell membrane fluidity [178], which would support substrate diffusion into the cell and thus the activity of the whole-cell catalyst. Despite the negative impact of organic cosolvents on the metabolism, viability and energy balance of the biocatalyst, efficient dissolution of the substrate is required for efficient catalysis. This was impressively demonstrated in a reaction using 1.0 mM substrate **1** without cosolvent, resulting in a decrease of the specific activity of the whole-cell catalyst by ca. 70 %.

Additionally, the impact of the tested solvent may be also influenced by a different solubilization behaviour of substrate **1**. Regarding the tested cosolvents, methanol has the highest relative polarity (0.762) compared to DMSO and isopropanol with relative polarities of 0.444 and 0.546, respectively [274]. It could be that at higher substrate concentration, depending on the used cosolvent, substrate **1** was not completely dissolved anymore in the reaction mixture, which would impact conversion and specific activity. A limited solubility of **1** would usually be observed by precipitate formation, however, since the whole-cell biocatalyst was used, the reaction mixture was turbid anyway. So, it cannot be concluded now if cosolvent addition was sufficient in each case

to keep the substrate fully dissolved. Further test would be necessary to investigate the solubility of **1** in different concentrations in reactions with addition of the three cosolvents tested herein.

Apart from the cell viability and integrity, the tested organic solvents will also affect the structure and stability of the enzymes used in our cascade. Direct information on the solvent tolerance of purified β-etherases and glutathione lyases is not available, however, some indications are found in literature [122]. Thus, it was shown that β-etherases are compatible with higher DMSO concentrations (up to 25 %) than glutathione lyase LigG-TD (up to 10 %). For future applications, further cosolvents could be tested in combination with our whole-cell catalyst and the impact of each solvent on the activity and stability of the involved enzymes should be investigated. Moreover, different methods have been developed over the years to "stabilize" a biocatalyst in the presence of organic solvents. Examples are the use of hydrophobic solvents in a two-phase system [179,180] or the immobilization of the catalyst [181,182], which could be tested for our system as well.

4.2.2 Glucose impact on GSH regeneration and whole-cell catalyst reuse

The whole-cell approach was developed to enable GSH provision and regeneration by the *E. coli* metabolism. As in each step of the tested biocascade (**Fig. 4.1**) one molecule of reduced GSH is consumed, two molecules of GSH per molecule of substrate **1** are required to achieve full conversion to the final product VG with simultaneous generation of GSSG. The latter is reduced again by *E. coli*-innate, NADPH-dependent glutathione reductase, according to equation **Eq. 5** [183].

$$GSSH + NADPH + H^+ \rightarrow 2GSH + NADP^+ \qquad (\boldsymbol{Eq.\,5})$$

Hence, incorporating our biocascade into the metabolism of *E. coli* could interfere with the central cell metabolism, especially pathways requiring NADPH as a cofactor.

In bacteria, NADPH is generated in the oxidative phase of the pentose phosphate pathway and can also be formed from $NADP^+$ by the use of NADH within a cell [275]. As glucose is the preferred carbon source for *E. coli* [275] and the initial substrate for the pentose phosphate pathway as well, it was assumed that the addition of glucose to the reaction may compensate for NADPH consumption. Moreover, the synthesis of GSH in the cell is dependent on ATP, which is mainly produced by ATP-synthase using the transmembrane proton motive force [275]. The latter is created by the electron transport chain driven by activated carriers such as NADH, which are produced via glycolysis and

TCA cycle starting directly from glucose. Moreover, ATP can also be generated by the Embden–Meyerhof–Parnas pathway directly from glucose [275]. Based on this information, we assumed that glucose addition should have a positive impact on the activity of the whole-cell catalyst. Surprisingly, no positive effect of glucose addition was observed in the whole-cell reaction. This was rather unexpected since in a similar cascade, expressing enantioselective pinoresinol reductase from *Arabidopsis thaliana* (AtPrR2) or enantiocomplementary pinoresidol lariciresinol reductase from *Forsythia intermedia* (FiPLR) for NADPH-dependent pinoresinol reduction while using the metabolism of *E. coli* C41 (DE3) to regenerate NADPH, addition of 0.11 M (20 g/L) glucose to the reaction mixture had a positive impact on the catalyzed whole-cell reaction [184]. It was later demonstrated that in that whole-cell reaction, another side reaction catalyzed by the *E. coli* metabolism was consuming cell-innate NADPH, causing a shift in the NADP+/NADPH equilibrium and thus, glucose addition indeed ensured higher NADPH concentration within the cell [184]. Our cascade, however, directly started from *rac*-**1** as the substrate and thus, no side reactions were expected or observed.

Obtained results in our experiment suggest that the required amounts of GSH and NADPH for conversion of 0.8 mM substrate via the whole-cell catalyst could be provided by the cell alone. Based on literature, *E. coli* cells contain glutathione in a concentration of 10 µmol [135] to 27 µmol per gram of dry cells [136]. Considering that our 1 mL reactions with OD_{600} of 40 contained about 13 mg of dry cell weight and that two molecules of GSH are required for cleavage of a β-*O*-4 aryl ether bond, the applied substrate concentration of 0.8 mM required at least 4.5-times higher GSH concentration compared to the one present within the cells. Hence, substantial GSH recycling and associated NADPH synthesis/regeneration obviously occurred within the cells without requirement for external glucose. Interestingly, even at increased substrate concentrations, no activity improvement was observed upon glucose addition, although in the reactions with 10 mM substrate (with and without glucose addition), only approx. 50 % conversion could be reached. This suggests that the amounts of cofactors NADPH and GSH produced by the cell metabolism are not the limiting factor in the reaction. Incomplete conversion in these cases may actually be caused by continuous whole-cell catalyst death [185]. Another possibility of the limited conversion in the high substrate concentration experiment may be explained by the formation of the final product, veratrylglycerone. This compound is not further consumed by the *E. coli* strain and is structurally similar to several compounds that are confirmed to have a toxic effect on the *E. coli* metabolism [186].

On the other hand, when the solvent concentration was dramatically increased to 15 % v/v and 25 % v/v, allowing for a better dissolution of the substrate in the reaction mixture

and thus more substrate to be available for the catalyst to convert, the addition of glucose resulted in a conversion improvement of 26 % and 22 %, respectively. In this case, however, the benefit of glucose was maybe rather to help the whole-cell catalyst protect against the solvent by providing carbon and energy for cell membrane regeneration processes, rather than to provide GSH and NADPH for the reaction. Similarly, when re-using the cells for multiple rounds of conversion of 0.8 mM substrate, glucose addition had a positive effect on the reaction rate as well. On average, the second, third and fourth rounds of reaction showed activity improvements by 16 %, 20 % and 10 %, respectively. It was also observed that in the reactions with added GLC, the reused cells were more stable and viable, meaning that the cell pellet after centrifugation was firmer and less cell lysis had occurred. Hence, it is assumed that glucose addition as carbon and energy source was beneficial for cell viability rather then GSH recycling.

Overall, by using the whole-cell catalyst, substrate **1** was successfully converted without the addition of the cofactor GSH required for the reaction. It is confirmed that the cell metabolism was able to recycle the formed GSSG to GSH for further rounds of substrate conversion. The whole-cell biocascade setup can theoretically be applied for a variety of substrates for which the β-etherases and glutathione lyases show activity. The only requirement is an uptake or diffusion of the substrate into the cell.

4.2.3 Kinetic resolution using the whole-cell catalyst

Since β-etherases display strict stereoselectivity, a possible application of our whole-cell approach was the provision of enantiomerically pure lignin model substrates, such as (*R*)-**1** or (*S*)-**1**. For this, three different setups were compared.

In the first approach, the whole-cell catalyst harbored either LigE or LigF-NA only. In this case, the catalyst was able to convert only one enantiomer of *rac*-**1** to form the respective glutathione adduct, while the non-reacted enantiomer remained unconverted. Recycling of the consumed GSH was not possible. Results of the kinetic resolution of **1** in this case did not achieve full conversion of one substrate enantiomer. Hence, it was assumed that an ether bond cleavage could only proceed until the cell-innate GSH was depleted. The absence of glutathione lyase in this construct prevented further degradation of the glutathione adduct and thus GSH recycling, resulting in incomplete kinetic resolution. The general function of GSTs is to maintain redox balance and form exportable conjugates from various toxins. In this set up, lack of the glutathione lyase caused formation of the GSH-conjugates, which are more soluble and thus more likely to be transported out of the cell. Several ATP-dependent transporters belonging to the family of multidrug resistance proteins (MRP) [187–189] ensuring transport of GSH-conjugates and

GSSG out of the cell have been described in eukaryotic organisms. The presence of such a transporter for a GSH-conjugate in *E. coli* has not been described but is assumed to be present as well.

In the second setup, *E. coli* cells harbouring a combination of one β-etherase and the glutathione lyase (LigE together with LigG-TD or LigF-NA together with LigG-TD) have been used. In this case, the formed glutathione adduct was further converted into VG, but also GSSG, which could be recycled. This setup gave satisfactory results as full kinetic resolution of **1** was achieved. In the subsequent semi-preparative-scale kinetic resolution of **1**, full conversion of the desired enantiomer could be achieved as well, as confirmed by chiral HPLC analysis, yielding 39 mg of enantiopure (*R*)-**1** and 43 mg of enantiopure (*S*)-**1**. Using whole-cell biocatalysis for kinetic resolutions has already been proven efficient previously [190]. Jeon *et al*. [190] reported that the co-expression of two enzymes; a stereoselective amine dehydrogenase (AmDH) and a cofactor-regenerating NADH oxidase (Nox), in *E. coli* BL21 (DE3) enabled the effective kinetic resolution of various racemic amins. Similar to the observations made in this thesis, their reported kinetic resolution reactions within the whole-cell catalyst yielded enantiopure amines with >99 % *ee* without any requirement for cofactor addition. In comparison, the kinetic resolution using purified enzymes gave the desired products with only 26 % *ee*. It was also shown that use of a whole-cell biocatalyst increased the stability of the heterologous enzymes and reduced substrate and product inhibition. Similar to our experiment, the whole-cell biocatalyst was further used in semipreparative scale, and 50 mM *rac*-2-aminoheptane and 20 mM *rac*-α-methylbenzylamine were successfully resolved to produce enantiomerically pure compounds with satisfactory yields [190].

The third approach tested in this thesis combined two separately produced whole-cell catalysts, one containing the β-etherase, the other containing the glutathione lyase. In order to obtain (*S*)-**1**, cells expressing LigE were used in combination with cells expressing LigG-TD. The (*R*)-**1** enantiomer was obtained by combining cells expressing LigF-NA with cells expressing LigG-TD. This setup, separating each reaction step of the cascade into different cells, yielded only an incomplete kinetic resolution, even after 30 h of reaction. This is likely explained by a hindered GSH regeneration and the requirement for different transport steps. Since production of the GSH-conjugate happens in one cell, whereas further conversion of this conjugate will take place in another cell, the produced GSH-adduct has to be transported from one to the other cell, which might not be efficient. Moreover, if the GSH-adduct is transported from one cell to another, GSH can only be recycled in the cells containing the glutathione lyase, and not those expressing the β-etherase. Regenerated GSH from the glutathione lyase-

containing cells could, however, be exported from the cytoplasm to the periplasm via the *CydDC* exporter [191] and further through the outer membrane to the reaction environment. From there, the β-etherase-containing cells could import the regenerated GSH via the *yilABCD* importer to the cytoplasm [192], where the GSH could be used for further rounds of **1** conversion. This transcellular transport is possible, but will take some time. Moreover, all of the mentioned transporters are ATP-dependent [135,191,192] and thus, more energy is required for sufficient GSH regeneration in the kinetic resolution of **1** using this setup.

Previously, Xiao *et al*. [193] also tested various setups for combination of (2*R*,3*R*)-2,3-butanediol dehydrogenase from *Bacillus subtilis* (ydjL) and NADH oxidase from *Lactobacillus brevis* in an enzymatic cascade for production of (3*R*)-acetoin, (3*S*)-acetoin and (2*S*,3*S*)-2,3-butanediol. Similar to our observations, the approach where the stereoselective dehydrogenase and the cofactor-regenerating NADH oxidase were expressed in separate cells, low conversion of *meso*-2,3-butanediol was observed. Only upon co-expression of both enzymes, (2*S*,3*S*)-2,3-butanediol with >99 % *ee* was produced. Even though in their case the cofactor to be regenerated was NAD^+ and not GSH, these results support our observation that co-expression of the cofactor-regenerating enzyme in the same cell as the cofactor-consuming enzyme is crucial in order to achieve efficient kinetic resolution as well as effective cofactor regeneration in whole-cell biocatalysis [193].

Overall, it was shown that only the co-expression of β-etherase and glutathione lyase within one whole-cell biocatalyst enabled efficient kinetic resolution of lignin model substrate **1**, presumably due to the possibility of efficient intracellular GSH recycling.

4.2 Mutational studies

4.2.1 LigE

The previously solved crystal structures of LigE and LigF [112] were very helpful in allowing us to gain more insight into the catalytic mechanism of both etherases and the role/involvement of active-site residues. The herein obtained results can now be the starting point for furture protein engineering of β-etherases, tailoring their catalytic performance and possibly enlarging the substrate spectrum.

The LigE structure was previously solved with the bound cofactor GSH, while a substrate-bound structure could not be obtained. Using Yasara, docking of the substrate (*R*)-**1** revealed a possible binding pocket in LigE. However, none of the predicted binding

modes appeared to represent a productive substrate binding conformation, which would also enable catalysis. For LigE, this has been reported before [111]. Also, recently performed MD (molecular dynamics) and QM/MM (quantum mechanics/molecular mechanics) simulations did not provide successful predictions due to the structure of the entrance to the catalytic region [116]. For the predicted S_N2 mechanism of β-etherases, the distance between the substrate's β-carbon and the GSH thiolate is expected to be lower than 4 Å [116]. However, from the calculated binding modes, the shortest C_β-to-S distance was 5.622 Å. Hence, the predicted substrate binding pose likely does not represent the productive substrate conformation. The difficulties to obtain a reliable substrate binding pose for LigE are likely explained by the enzyme's crystal structure. It has been proposed that the active site of LigE is partially occupied by aromatic residues Tyr23 and Trp107 in this structure [116]. Additionally, the crystal structure of LigE was only obtained for a C-terminally truncated variant (removal of 27 residues at the C-terminus of LigE) and thus, the structure of this truncated LigE variant might not reflect the catalytically active conformation of LigE [112]. Based on our results and literature data [116], it is anticipated that for substrate (*R*)-**1** to bind to the active site, extensive conformational changes in the structure of LigE have to occur. The herein presented substrate binding analysis of **1** using tryptophan fluorescence, indicating cooperative substrate binding behavior, confirms the presence of structural changes upon substrate binding.

Molecular docking represents a bioinformatic modelling tool that attempts to predict the structure of a ligand-receptor complex (enzyme-substrate, drug-protein, drug-nucleic acid, protein-nucleic acid, and protein-protein) with optimized conformation and the intention of possessing less binding free energy [194,195]. During the calculation of molecular docking, thousands of possible ligand/receptor poses for the interaction are investigated and evaluated. Based on the ligand/receptor structures, the pose with the lowest energy score is predicted as the "best match", referring to the binding mode. Despite the continuous development and widespread use of molecular docking, the process of rapid and accurate prediction of ligand/receptor interaction is complex and challenging, and drawbacks continue to affect the calculated outcome [196,197]. Knowing the structural location of the binding site, as well as the flexibility of ligand and/or receptor, the charge of the structures' environment and the interactions with the adjacent water molecules can complicate the quantitative analysis and thus the "correct" prediction of the binding pose [195,198]. Up to now, several studies comparing different docking programs have been published [199–202] and it was shown that molecular docking methods average an approx. 60-75 % success rate on the identification of correct poses [202,203]. The docking program AutoDock Vina that was used for the molecular docking presented in this thesis

(as an extension of the virtual reality software YASARA), exhibits up to 78 % of accuracy for the receptor-ligand docking. This makes the AutoDock Vina one of the more efficient, accurate and fast docking programs available [202,204].

Despite the unsatisfactory docking results with the crystal structure of LigE, several amino acid residues in proximity to the substrate and GSH were chosen to be exchanged by alanine (alanine scanning). All amino acid exchanges had a severe impact on the activity of LigE towards both tested model substrates **1** and **3**. Selected amino acid residues Tyr23, Trp107, Phe115, Tyr122, Phe142 and Trp197 of LigE were compared to a multiple sequence alignment (MSA) of LigE-homologs (shown in chapter 7.9), which consisted of 37 sequences acquired from a BLAST search. According to this MSA, positions Trp107 and Tyr122 seem to be conserved among all compared sequences, which confirms the findings published by Voß *et al*. [111]. Positions Tyr23 and Trp197 are more flexible as also phenylalanine is found at these positions in the homologous sequences. This, however, represents only a minor alteration since the residues are all bulky, hydrophobic and aromatic, which is in direct agreement with previously published data [111]. On the other hand, the positions Phe115 and Phe142 of LigE seem to have a relatively random amino acid residue occupancy in the MSA.

Interestingly, despite the fact that no positive effect on the activity of tested mutants was observed, the impact of individual mutations on specific activity was substrate-dependent. Mutation Y23A resulted in a dramatic loss of activity (~98 %) towards substrate **1**, while the same mutant displayed only an activity reduction by ~60 % towards substrate **3**. In contrast, the mutant F142A exhibited decreased activity by ~70 % towards substrate **1**, compared to 98.5 % towards substrate **3**. These results suggest that the different model substrates bind slightly differently in the enzyme's active site. Hence, active site residues are more or less important for substrate binding depending on the substrate structure. A strong substrate specificity of LigE was previously reported [110], as the specific activity of the enzyme was greatly dependent on the substitution pattern at the phenyl ring(s). This confirms the fact that mutations that have a severe impact on one substrate are not confined to have the same effect on another substrate [205].

Until now, several structural analyses of LigE have been published [111,112,115,116]. By solving the x-ray crystal structure of the LigE-GSH complex (LigE as C-terminally truncated version), Helmich *et al*. [112] identified the amino acid residues responsible for stabilization of GSH in its binding pocket (Asp71, Ser72, Val59, Arg138 and Tyr133). From these, residue Arg138 was subjected to an alanine mutation in this thesis. This revealed the importance of Arg138 for catalysis since its exchange by alanine led to a complete loss of activity. Hence, proper GSH binding and stabilization in the active site

seems to be crucial for LigE catalysis.

Similarly, it was proposed [112] that amino acid Ser21 may be of catalytic importance, as the hydroxyl group of serine could activate the GSH by providing a hydrogen bond to the sulfhydryl group of GSH, promoting formation of the thiolate anion [94,206]. This hypothesis, however, could not really be confirmed since the mutant S21A did not completely lose activity, although a severe activity loss was detected [112]. Despite the fact that Ser21 does not seem to be absolutely crucial for LigE activity, only alanine and glycine at position 21 were found in this thesis to retain some catalytic activity of the respective mutants. The exchange to alanine resulted in a great loss of activity (>96 %), while the decrease in activity of mutant S21G was less severe (70-80 %, depending on the substrate). Hence, it seems to be important for activity that the amino acid at position 21 carries only a small side chain. In the MSA consisting of 37 sequences of LigE homologs (chapter 7.9), position 21 is strictly conserved, thus supporting the theory that Ser21 is important for LigE catalysis by supporting GSH binding or ensuring proper thiol orientation [112].

In the presented mutational studies of the active site of LigE, we were able to confirm the non-essential role but importance of residue Ser21 for LigE catalysis. Additionally, residue Arg138 was found to be essential for GSH binding. However, no further residues of catalytic importance could be identified among the mutated positions. Hence, further studies (computational and experimental) of LigE are necessary in the future to further expand our knowledge on the catalytic mechanism of this β-etherase.

4.2.2 LigF

A comparable approach was followed for the analysis of the LigF active site. Based on the previously solved crystal structure of LigF containing the cofactor GSH, substrate (*S*)-**1** was docked in to identify potentially important active-site residues. This time, the predicted binding pose likely represented the productive binding mode, as concluded based on substrate geometry and the distance between the β-carbon and the GSH thiolate (3.421 Å). It also agreed with the recently published substrate binding mode resulting from MD and QM/MM simulations of LigF [116]. Based on the calculated prediction, several amino acid residues located in proximity to the docked substrate and GSH were targeted by alanine scanning: Phe7, Asn12, Cys107, Trp108, Val110, Ser111, Trp115, Ile119, Ile122, Gln144, Lys147, Trp148 and Ile199. Moreover, residue Pro142 was selected even though it is located furthest from the GSH (8.8 Å). When comparing these positions to the corresponding residues of homologous LigF sequences in a multiple sequence alignment (MSA) consisting of 41 sequences obtained by BLAST analysis (shown in chapter 7.10), strict conservation was detected for residues Phe7,

Asn12, Cys107, Trp108, Val110, Ser111, Trp115, Gln144, Lys147, Trp148 and Ile199. The large number of conserved amino acid residues among LigF homologs in the active site indicates that these residues could have important catalytic or structural functions. A high level of conservation in sequence clusters spanning the active sites of LigF-type β-etherases was recently reported [111], implying that substantial catalytic function is encoded in these regions and has been maintained by natural selection during biological evolution. This conservation of sequence motifs among homologous enzymes was recently used to broaden the scope of available β-etherases [111]. A peptide pattern recognition (PPR) tool [207] was used to classify and cluster LigE- and LigF-type β-etherases form a large number of homologous sequences by the presence of specific peptide motifs. Three motifs, LYS**F**GPxA**NS**xKP, TESTVICEYLEDxxP and AxMRxWTKWVDEYF**CW**C**VS**TxG**W**, were identified for LigF-type β-etherases [111]. Several amino acid residues mutated in this thesis correspond with residues located in the reported sequence motifs (F7, N12, S13, C107, W108, V110, S111, W115; marked bold in previous motifs). In contrast, residues I119 and I122 are less conserved because also valine and methionine were found at these positions in the MSA (chapter 7.10). Alanine exchange for all mentioned amino acid residues in LigF resulted in a severe impact on enzyme activity; in case of W115A and W148A, even a complete loss of activity was observed.

Furthermore, special attention was drawn to position Pro142, for which the alanine exchange resulted in improved activity with both of the tested substrates. Using site saturation mutagenesis, several amino acids at position 142 led to improved activity compared to the wild-type enzyme. Among them serine, tyrosine, phenylalanine, histidine, as well as the previously mentioned alanine, which represents a rather diverse set of amino acids regarding their characteristics. Additionally, according to the MSA of LigF homologs (chapter 7.10), proline at position 142 is not conserved, as multiple amino acids with variable properties (proline, valine, isoleucine, alanine, phenylalanine and lysine) are found at this position in the MSA. This large natural diversity of position 142 speaks against a catalytic function of this residue. Interestingly, position Pro142 within the enzyme's structure is located rather far from the predicted active site and hence, its role for enzyme activity cannot be fully understood. It might be, however, that position 142 is important for protein dynamics of LigF, which could play a role during catalysis. It has also been shown for other enzymes that single mutations located remotely from the active site (>10 Å) can have a significant impact on catalytic function [208–214]. Moreover, LigF was subjected to random mutagenesis by error-prone PCR very recently and two single mutations (P86S and D128G) were found to exhibit improved specific activity [215].

Both residues are again distantly located from the active site (both more than 20 Å, based on our docking). As mentioned, those positions as well as P142 could have an impact on structural and dynamic features of the enzyme, and therewith also enzyme activity, which might imply larger protein motions during catalysis. However, further studies are needed to confirm the biological relevance of position P142 for LigF catalysis.

Additionally, site-saturation of Trp115 revealed the importance of the aromatic character of this residue for substrate binding. A complete loss of activity was observed when Trp115 was exchanged by non-aromatic amino acids, whereas LigF mutants W115F, W115T and W115H still displayed activity. However, even substitution of W115 with these aromatic residues (phenylalanine, tyrosine and histidine) had a severe negative impact on activity and enzyme kinetics. The required aromatic character of this residue, which is located in direct proximity to the docked substrate, is very likely responsible for the correct positioning of the substrate in the active site.

The importance of Trp148 and Gln39 for LigF catalysis was confirmed by the observed complete loss of activity after an exchange by alanine. These amino acid residues were previously described to interact with the glycine part of GSH. Hence, our results suggest that the correct positioning of GSH is crucial for LigF catalysis. Binding of the GSH cofactor to the G-site of the thioredoxin domain is stabilized by a network of hydrogen bonds provided by surrounding residues. Apart from W148 and Q39, these include also E65, S66, Q52, Q144 and H40 [112,216].

As catalysis by GSTs typically depends on activation of the GSH molecule to form its thiolate form via hydrogen bonds with active-site residues [217], it was previously suggested by Helmich *et al*. [112] that the hydroxyl group of Ser13, which is located in the active site of LigF, could hold this function. Further research, however, refuted this theory, since no complete loss of activity was observed when Ser13 was exchanged to alanine [112]. Site-saturation mutagenesis of this residue in this work revealed several amino acid residues (alanine, glycine, cysteine, asparagine) that still ensured catalytic activity, though with severe activity losses. This suggests that Ser13 is indeed important but not essential for LigF catalysis. Later, Kontur *et al*. [115] suggested that activation of GSH by deprotonation of the thiol may involve the conserved residue Asn12. Site-saturation mutagenesis of this residue in the present work revealed that position Asn12 alone is also not crucial and thus not solely responsible for GSH activation. Different amino acid residues (alanine, histidine, methionine, phenylalanine, arginine, threonine, lysine) at position 12 were shown to maintain catalytic activity. Moreover, mutation N12H even had a positive effect on enzyme activity. This is likely explained by the fact that histidine is an even better base for proton abstraction. Histidine is often found in enzymes

where de- and reprotonation steps are required during catalysis [218,219]. Based on MD and QM/MM simulations of LigF catalysis, Prates *et al*. [116] recently proposed that both amino acid residues, Asn12 and Ser13, are cooperatively responsible for activation of GSH. Asn12 interacts directly with the GSH thiolate via a hydrogen bond, while Ser13 hydrogen bonds with the thiolate via a water molecule as shown in **Fig. 4.2**. The experimental data collected in this thesis support the theoretical data published by Prates *et al*. [116]. While single mutants, in which Asn12 or Ser13 were exchanged for non-hydrogen bonding residues, enzymatic activity, activity was not completely lost. For the corresponding double mutant LigF-N12A+S13A, however, no residual activity could be detected anymore. This experiment proves the importance of residues Asn12 and Ser13 for GSH activation, and their cooperative effect on LigF catalysis, thus supporting the theoretical data previously published.

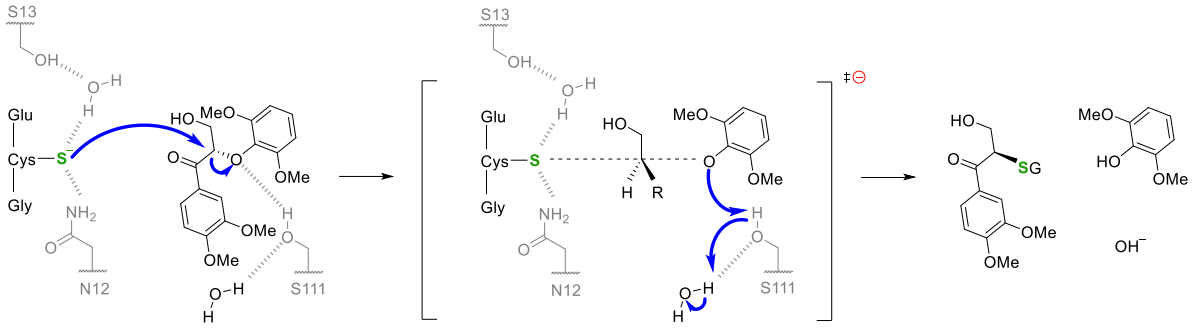


Fig. 4.2: Proposed S_N2 reaction mechanism of LigF in cleavage of β-*O*-4 aryl ether bond in substrate **1**. Blue arrows indicate electron flow and dashed lines represent important hydrogen bonds for catalysis [116].

Furthermore, in the LigF mechanism proposed by Prates *et al*. [116], residue Ser111 is responsible for mediating proton transfer from a water molecule to the oxygen of the leaving group (2,6-MP in **Fig. 4.2**) upon ether bond cleavage. The importance of this residue was confirmed by mutation S111A, which resulted in a severe activity loss (≥98.5 %). Enzymatic activity of LigF S111A, however, was not abolished completely, suggesting a possible compensation by a different residue or water directly. This could possibly be ensured by neighbouring Trp115 (H-to-O distance of 2.566 Å) or by a water-mediated proton, but further analysis is necessary to confirm these theories.

In summary, with the herein presented mutational study of the catalytic site of LigF, the cooperative importance of residues Ser13 and Asn12 for GSH-activation was demonstrated. Moreover, residue Trp115 was shown to play an important role for LigF catalysis, likely for correct substrate positioning as only aromatic residues at this position retained residual activity. Additionally, different LigF mutants with improved catalytic activity have been identified, among them Asn12His as well as several mutants at

position Pro142, implying that the catalytic properties of LigF can be further improved by protein engineering and that structural changes of the enzyme may occur during catalysis.

4.3 Substrate binding studies

Tryptophan is an α-amino acid that is relatively rare in proteins. On average, only ~1 % of the amino acid residues in protein structures is reported to be tryptophan [220,221] which is also reflected by the fact that there is only one DNA codon for tryptophan. This has been concluded to be connected with evolutionary and functional reasons since tryptophan is the largest amino acid residue and its synthesis requires more ATP than any other proteinogenic amino acid. Therefore, it is also assumed that, in most cases, tryptophan plays an important biological and functional role in proteins. Moreover, as an aromatic amino acid, tryptophan possesses intrinsic protein fluorescence, which is strongly affected by its local environment. This property has been used in protein-related research to quantify conformational changes of tryptophan residues as well as changes in the tryptophan environment.

The LigE monomer contains 11 tryptophan residues in its structure. Six of them (W25, W114, W136, W141, W178, and W216) are buried inside the hydrophobic core of the protein, one (W68) is located at its surface and two (W73 and W105) are located in the dimer interface. The latter are likely involved in the interaction of both monomers. The remaining two tryptophan residues, W107 and W197, are found in the active site (**Fig. 4.3A**). Hence, the intrinsic fluorescence of these tryptophan residues should change upon binding of the substrate. Similarly, one LigF monomer contains eight tryptophan residues in its structure, two of which (W43 and W57) are located at the surface. Another two (W98 and W101) are located at the dimer interface and are believed to play a role in dimer formation. Residue W225 is located inside the protein's hydrophobic core structure, while W148 was shown to interact with GSH. Again, two residues, W108 and W115, are located in the active site and their fluorescence is expected to change upon substrate binding (**Fig. 4.3B**).

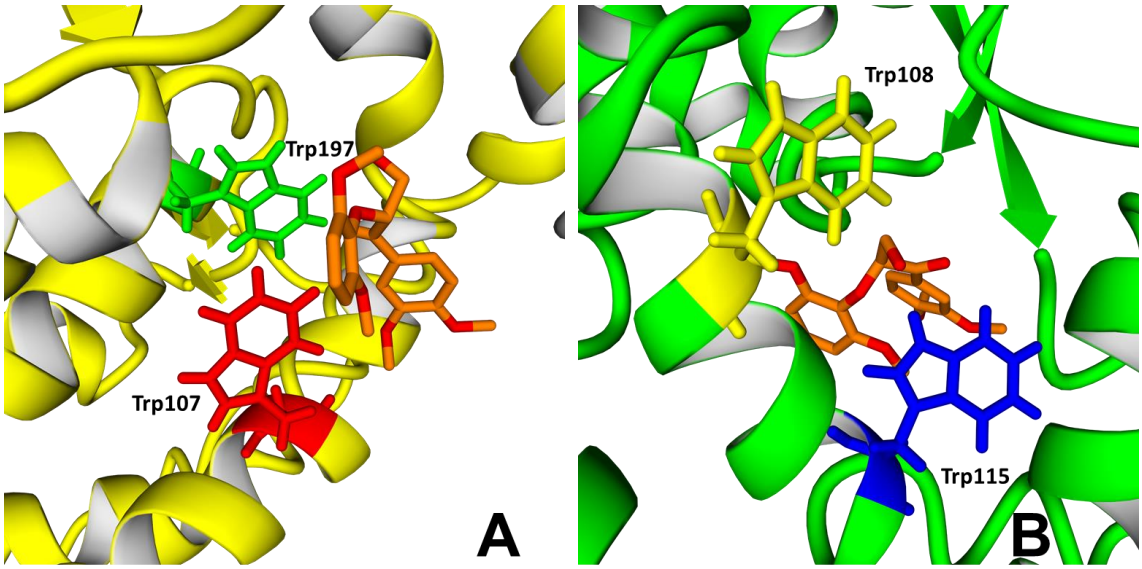


Fig. 4.3: (**A**) Active site of LigE (yellow, PDB code: 4YAN) with the docked substrate β(*R*)-2,6-MP-VG (orange) and the tryptophan residues located in the active site: Trp107 (red) and Trp197 (green), which intrinsic fluorescence is expected to be altered upon substrate binding. (**B**) Active site of LigF (green, PDB code: 4XT0) with the docked substrate β(*S*)-2,6-MP-VG (orange) and the tryptophan residues located in the active site: Trp108 (yellow) and Trp115 (blue), which intrinsic fluorescence is expected to be altered upon substrate binding.

Thus, a tryptophan fluorescence assay was developed and applied in this thesis to analyze the binding of different model substrates in LigE and LigF. This revealed that in addition to lignin model substrates **1** and **2**, also substrate **4**, containing a thioether bond, displays high affinity for binding to the active sites of LigE and LigF despite the fact that this substrate is not converted by any of the tested β-etherases. According to collected data, substrate **4** can bind to the active sites of LigE and LigF, even though it is not actively converted by these enzymes. The latter is likely explained by the fact that the cleavage of the thioether bond by nucleophilic attack of GSH was found to be endothermic according to DFT calculations. Hence, compound **4** could be useful as a substrate analog in enzyme crystallography or mechanistic studies of β-etherases in the future. Moreover, compound **5**, lacking the β-*O*-4 aryl ether bond, resulted in changes of intrinsic tryptophan florescence upon titration to LigE and LigF. However, these changes did not follow any trend. This might be explained by a random interaction of this compound with tryptophan residues located on the surface of the enzymes as well as in the active sites. This data suggests that substrate **5** does not effectively bind to the active site of enzymes LigE and LigF. Binding of substrate **5** to the active site of LigF-NA enzyme was previously analyzed by Voß [269] using micro scale thermophoresis (MST). The received results of the measurement were unfortunately not unambiguous and it was not possible to conclude if real binding or only binding artefacts were measured. Additionally, no binding could be observed for substrate **6**, which represents a model

substrate for the NAD-dependent Cα dehydrogenases (LigD, LigO, LigN and LigL) that are described as early enzymes in the lignin β-*O*-4 aryl ether cleavage pathway in *Sphingobium* sp. SYK-6 [105,118]. In this pathway, enzymes LigD, LigO, LigN and LigL catalyze the stereospecific oxidation of the Cα hydroxy group of **6** to generate a keto group in Cα position [102,222], which is a prerequisite for ether bond cleavage by β-etherases [107,110,119]. This keto group is suggested to be essential for subsequent cleavage due to intramolecular activation of the ether bond and the altered molecule geometry. Our experimental data shows that **6** is also not bound to the active site of β-etherases.

Furthermore, the binding of substrates **1**, **2** and **4** was described by sigmoidal binding curves, suggesting cooperative binding. Cooperative substrate binding, as described by the Hill equation, can be observed in some multimeric enzymes where binding of one substrate molecule alters the affinity of another binding site due to ligand-induced conformational changes [223–225]. In the case of LigE, LigF and their mutants, Hill coefficients close to 2 were observed, indicating significant cooperativity, as both β-etherases form homodimers. This cooperativity, however, was not observed during kinetic measurements. The reason for this difference is not clear yet. It cannot be excluded that the use of a racemic substrate mixture, of which only one enantiomer is actually converted by the enzyme, or the conditions of the tryptophan fluorescence assay itself are influencing the binding behavior. Hence, further investigation is necessary to answer the question if substrate binding to LigE and LigF happens indeed in a cooperative fashion or not. In general, GSTs were for a long time considered as noncooperative enzymes, meaning that substrate binding to one subunit of the homodimeric protein would not modify the catalytic properties of the other one [226–228]. However, mutagenesis studies of the homodimeric maize glutathione S-transferase have shown a positive cooperativity for 1-chloro-2,4-dinitrobenzene (CDNB) binding as well as for binding of 1-hydroxyl-2,4-dinitrobenzene (HDNB) [229]. Positive cooperativity in this case was induced by a mutation on the dimeric interface and was explained by causing conformational communication between monomeric subunits upon substrate binding. Additionally, positive cooperativity for GSH binding was confirmed in mutational studies of human GST P1-1 [230], which is responsible for the conjugation of prostaglandins J2, A2 [231] and halide ions [232]. In addition, wild-type human GST P1-1 was reported to display temperature-dependent positive and negative cooperativity of GSH binding above 35˚C and below 25˚C, respectively [233].

Additionally, it was shown that LigE and LigF mutants with severely reduced or even completely abolished activity still displayed substrate binding according to the tryptophan fluorescence assay. These results suggest that the single mutations in either LigE or

LigF, independent from their impact on catalytic efficiency, did not completely prevent substrate binding but rather resulted in unfavorable positioning of the substrate with regard to catalysis. Moreover, binding of substrate **1** to the active site of LigE and LigF was highly affected when the cofactor GSH was present. In this case, binding affinities were generally 2.4-2.8-times lower compared to measurements without GSH. This drop in affinity could simply be the result of iterative lignin model substrate conversion and product release during the tryptophan fluorescence assay due to the presence of GSH. On the other hand, it cannot be ruled out that the enzymes' affinity for substrate binding is indeed affected by the cofactor GSH. The latter could actually be tested in the future by using compound **4** as model substrate together with GSH in the tryptophan fluorescence assay, as compound **4** does bind to the active site of LigE and LigF but cannot be converted.

Furthermore, ITC was used to determine the binding affinity of LigF for GSH in the absence and presence of substrate. Results indicated strong binding of GSH to LigF (K_D = 80 µM) in the absence of substrate. This value is rather similar to the binding affinity of the glutathione S-transferase from *Plasmodium falciparum* for GSH (K_D = 70 µM) [234]. When model compound **4** was present, the K_D value increased, suggesting a lower affinity of LigF for GSH. In this case, however, the observed drop in affinity cannot be caused by substrate conversion, since model compound **4** cannot be cleaved by LigF. Hence, GSH binding is indeed altered in the presence of substrate and therefore, it is likely that also substrate binding is affected by GSH, as observed in the tryptophan fluorescence assay. Similar findings were previously observed for plant GST from *Mangifera indica* [235] for which crystal structures were solved for the GSH-enzyme and GSX-enzyme (GSX: model of the GSH-hexyl conjugate) complexes. Here, it was confirmed that binding of GSH to the enzyme active site at the G-site causes conformational changes to the enzyme. It was even suggested that the binding of GSH is required for proper structural formation of the substrate-binding site for catalysis [235]. This would also explain why substrate binding affinity of LigE and LigF wild-type enzymes and their mutants, as measured by tryptophan fluorescence assay, changed upon GSH addition.

Furthermore, the higher affinity of LigF for GSH binding in absence of substrate may have a physiological purpose. This way, release of GSH from the enzyme active site could be prevented before substrate is present for catalysis.

Furthermore, during ITC measurements a change in the value of N from 1 to 2 was observed upon substrate addition. The parameter N usually represents the number of ligand binding sites in the protein macromolecule. Hence, the obtained results could

suggest that there are two GSH binding sites per LigF monomer when 100 µM and 200 µM of model compound **4** was present, in comparison to only one binding site when only GSH was present. This is, however, rather unlikely based on the known LigF crystal structure and based on the shape of the binding curve (if two binding sites were present, the binding curve should have double-sigmoidal shape). Apart from the number of binding sites, the parameter N can be influenced as well by structural changes in the binding sites. In case of homodimeric LigF this could mean that upon substrate binding both GSH binding sites are structurally not identical anymore. This could explain why in our measurements without model compound **4** (and with only 50 µM of **4**) an N value of 1 is obtained, as here both GSH binding sites of the homodimer are likely symmetric, whereas this value increases to 2 as the concentration of **4** reaches the level of the enzyme concentration, which results in substrate binding to the enzyme. This binding of substrate analog **4** to the active site of LigF likely involves structural changes in the active site and hence, possibly also in the GSH binding site, which could make both GSH binding sites of homodimeric LigF asymmetric. Even minor alterations in the GSH binding site would be sufficient to cause changes in the N-value. Overall, our results support the hypothesis that structural changes occur in LigF (and likely also other β-etherases) upon substrate or GSH binding, a fact that has already been observed in multiple other GSTs [234–236]. Additionally, the molecular dynamics (MD) simulation of LigF by Prates *et al*. [116] in the presence of substrate and GSH indicated that 51 % of the predicted substrate positions involved hydrogen bonding between substrate and GSH. This result further supports our findings that binding of GSH and substrate influence each other. On the other hand, it cannot be completely ruled out that the observed increase in the N value during ITC measurements is simply an artefact. N is rather sensitive parameter in ITC as small inconsistencies in the concentrations of ligand or protein during the measurement may cause alterations in this value. Hence, further analyses are necessary in the future to better understand binding of substrate and GSH to the β-etherases and the possible involvement of conformational changes during binding.

Substrate binding studies with LigE and LigF enzymes using the tryptophan fluorescence assay provided valuable information about the binding behavior with multiple substrates as well as GSH. This analysis indicated that substrate binding is significantly affected in the presence of cofactor GSH. Likewise, ITC measurements revealed that the high affinity of LigF for GSH is also altered in the presence of substrate.

4.4 Novel non-lignin substrates of β-etherases

To investigate the biocatalytic potential of β-etherases further, they were tested in reactions with potential novel, non-lignin-like substrates. All of the tested compounds (**30**-**34**), provided by Prof. Dr. Dieter E. Kaufmann from the Institute of Organic Chemistry at TU Clausthal, Germany, contained a common key building block. This unit, known as bisphenol A (BPA, **Fig. 4.4A**), is an industrial chemical that has been commonly used in the synthesis of certain plastics and epoxy resins [237]. However, despite the positive effect of the compound on the properties of the resulting polymers (**Fig. 4.4B**), the use of BPA is very controversial due to its toxicity towards algae, invertebrates, fish, as well as humans [238,239]. Hence, the use of BPA is currently strictly controlled in European Union [276].

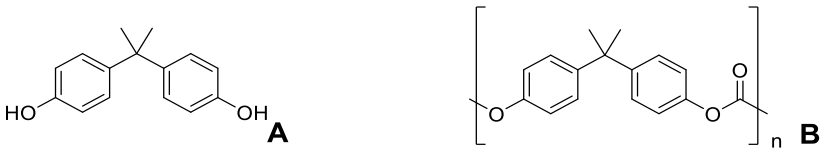


Fig. 4.4: Structure of the bisphenol A (**A**) and the commonly used bisphenol A-polycarbonate (**B**).

The toxic effects of bisphenol A are associated with early sexual maturation, altered behavior, and effects on prostate and mammary glands [240–242]. It has also been connected to cardiovascular disease, diabetes and male sexual dysfunction [243–248]. The European Union listed BPA as a potential endocrine disruptor and restricted its use in certain food-contact materials. It has been in use since the 1960's until today and released into the environment at low doses, although BPA and its polymeric forms are barely degraded in wastewater treatment plants [249]. Hence, methods are required to improve their degradability. As an alternative to conventional processes for phenol removal, several microbial approaches have been described to degrade a monomeric BPA, including ligninolytic enzymes from fungi (LiP, MnP, and laccase) [250–252] as well as bacterial strains *Sphingomonas* sp. BP-7 and *Sphingomonas yanoikuyae* BP-11R [253].
The compounds that were tested as β-etherase substrates within this thesis contained aryl ether bonds which were previously found to improve the stability of BPA polymers towards hydrolytic degradation and photodegradation in sunlight [254]. A low but proven catalytic activity of the bacterial β-etherases towards polyphenolic compounds **30**, **31**, and **32** was observed, demonstrating their potential for BPA degradation. The cleavage of aryl ether bonds (compound **30**) in non-lignin-like compounds by β-etherases as well as their activity for epoxide ring opening with GSH (compounds **31** and **32**) have not been described before. On the other hand, several strains of *Sphingomonas* sp. have

been reported to grow on monomeric BPA as the sole carbon source [253,255–259]. So far, the most promising strain appears to be *Sphingomonas* sp. strain AO1 [255], which was able to degrade 115 µg/ml of monomeric BPA within 6 hours. Recently, in a quantitative proteomics and metabolomics study of the BPA-degrading strain *Sphingobium* sp. BiD32 [258], which had been isolated from activated sludge, a hydroxylated metabolite has been detected and a *p*-hydroxybenzoate hydroxylase enzyme was identified to be involved in BPA degradation. Despite multiple studies of BPA-degrading bacterial strains, our knowledge of metabolic pathways for BPA degradation as well as the knowledge of involved enzymes and genes is still rather limited. What has been found out is that when multiple organisms were exposed to increased concentrations of BPA (or its alternative bisphenol F and bisphenol S), the activity and expression levels of GSTs were significantly increased [260–262]. Based on that, it is likely that GSTs (including β-etherases) might have potential for degradation of BPA and its analogs. Overall, the herein obtained data is a good starting point for further analysis regarding the application of β-etherases on non-lignin-like substrates.

5 Conclusions and future perspectives

Glutathione-dependent β-etherases and glutathione lyases catalyze the reductive cleavage of β-*O*-4 aryl ether linkages present in lignin. Due to their non-radical mechanism, these enzymes are promising biocatalysts for the selective depolymerization of lignin polymer as well as oligomers, and hence the production of aromatic platform chemicals from renewable sources. Lignin, as part of lignocellulose, is one of the most abundant aromatic polymers present on earth. Hence, its valorization is regarded as a valuable alternative to the synthesis of aromatics from mineral oil. β-Etherases and glutathione lyases belong to the superfamily of glutathione transferases and require 2 equivalents of the cofactor glutathione (GSH) for cleavage of each β-*O*-4 aryl ether bond.

In this thesis, an *E. coli*-based whole-cell biocatalyst was generated harboring glutathione-dependent β-etherases together with glutathione lyase for efficient supply of required GSH by the cells during β-*O*-4 aryl ether bond cleavage. By combining a stereoselective β-etherase and a less selective glutathione lyase, this whole-cell cascade enabled the kinetic resolution of 10 mM racemic lignin model substrate on a semi-preparative scale without the need to add expensive glutathione cofactor. Thus, the presented whole-cell system can be applied for the efficient preparation of enantiopure lignin model substrates containing β-*O*-4 aryl ether linkages. These molecules are useful substrates for mechanistic or kinetic studies of β-etherases and other ligninolytic enzymes. Similarly, chiral glutathione adducts formed by β-etherase catalysis can also be accessed using this whole-cell approach if no glutathione lyase is present. Moreover, the application of this intracellular GSH supply and recycling concept based on *E. coli* is not limited to β-etherases and glutathione lyases. In principle, it can be extended to other GSH-dependent enzymes. However, as the glutathione cofactor is provided within the cell, only substrates that are able to enter the cell can be converted with this approach. Examples could be the synthesis of enantiomerically pure aromatic compounds with a valuable pharmaceutical application like phenylpropanoids exhibiting anti Alzheimer's disease activity [263], compounds that promote the formation of new blood vessels [264], or neuroprotective compounds that have only been isolated from their natural sources so far [265]. Moreover, larger scale reactions in a fully-controlled environment could be explored. Furthermore, as the used organic cosolvents for substrate dissolution all exhibited a negative effect on the whole-cell catalyst performance, novel solvents that may also be less damaging to the catalyst and environment, such as solvate ionic liquids [266], should be tested.

For successful industrial application of any biocatalyst, extended knowledge of the catalyst's mechanism is necessary. In this thesis, the best studied β-etherases, LigE and LigF, underwent a detailed analysis of their active sites based on their previously solved crystal structures in combination with substrate docking. Mutational studies of the active-site residues of LigE and LigF revealed amino acid residues that are important or even essential for substrate binding and catalysis. Moreover, mutants with improved catalytic activity have been identified, establishing a good starting point for further protein engineering of β-etherases. Protein engineering will be necessary to improve the enzyme's properties, such as specific activity, thermostability, pH sensitivity and substrate specificity, for industrial application. Since the mutation of a remote residue, Pro142, located outside of LigF's active site, resulted in a boost of enzyme activity, LigF appears to be a good candidate for further activity improvement by random mutagenesis. Additionally, the importance and function of several amino acid residues for the catalytic mechanism of LigF have been proposed. Nevertheless, further analysis of active-site residues is still necessary, especially regarding Ser111 and the possible role of surrounding residues which are or might be involved in proton transfer to the leaving group during catalysis.

Results gathered in the mutational study of β-etherase LigE are, unfortunately, less meaningful as the productive substrate binding orientation could not be elucidated with the crystal structure of truncated LigE. This should be taken as a scientific challenge and further analyses, computational and experimental, should be conducted to reveal the substrate-bound structure of this promising enzyme. In that sense, a close cooperation between computational scientists and wet-lab researchers will be crucial in the future for efficient advancement of β-etherase-catalyzed lignin depolymerization and biocatalysis in general.

In connection to this, by conversion of BPA-related compounds, the substrate scope of β-etherases was demonstrated in this thesis to go beyond lignin-derived ether substrates. This opens new possibilities for potential application of these enzymes in waste water treatment and the degradation of non-lignin-like polymers. However, the data presented here is just the beginning and further investigations are necessary to uncover the full biocatalytic potential of β-etherases.

As indicated in this thesis, a collaboration of experts from various fields, including biochemical engineering, process engineering, bioinformatics, polymer science and biochemistry, will be necessary for the development of processes in which heteropolymeric lignin can be valorized in an efficient but environmentally sustainable way to replace petrol-based chemistry. The current climate situation, characterized by

increasing sea levels, melting icebergs, extreme droughts followed by extreme floods, increased wildfires, declining fresh water sources, and a rapid increase in Earth's average surface temperature should force not only the scientific, but also the layman community to do anything possible to reverse these changes and preserve the planet and its environment for future generations.

6 References

1. Haines A, Ebi K., *N Engl J Med* **2019**, *3*, 263–273.
2. Vijayavenkataraman S, Iniyan S, Goic R., *Renew Sustain Energy Rev* **2012**, *16*, 878–897.
3. Al-Ghussain L., *Environ Prog Sustain Energy* **2019**; *38*, 13–21.
4. Bugg TDH, Ahmad M., Hardiman EM, Rahmanpour R., *Nat Prod Rep* **2011**, *28*, 1883–1896.
5. Bugg TDH, Rahmanpour R., *Curr Opin Chem Biol* **2015**, *29*, 10–17.
6. Rinaldi R, Jastrzebski R, Clough MT, Ralph J, Kennema M, Bruijnincx PCA, Weckhuysen BM., *Angew Chem Int Ed* **2016**, *55*, 8164–8215.
7. Hendriks ATWM, Zeeman G., *Bioresour Technol* **2009**, *100*, 10–18.
8. Laureano-Perez L, Teymouri F, Alizadeh H, Dale BE., *Biotechnol Fuels Chem* **2009**, *121*, 1081–1099.
9. Wijffels RH, Barten H, Reith RH., *Bio-methane & Bio-hydrogen* **2003**, ISBN: 90-9017165-7
10. Zakzeski J, Bruijnincx PCA, Jongerius AL, Weckhuysen BM., *Chem Rev* **2010**, *110*, 3552–3599.
11. Davin LB, Lewis NG., *Curr Opin Biotechnol* **2005**,*16*, 407–415.
12. Boerjan W, Ralph J, Baucher M., *Annu Rev Plant Biol* **2003**, *54*, 519–546.
13. Davis R, Tao L, Tan ECD, Biddy MJ, Beckham GT, Scarlata C, Jacobson J, Cafferty K, Ross J, Lukas J, Knorr D, Schoen P., *Natl Renew Energy Lab* **2013**, *147*
14. Adler E., *Ind Eng Chem* **1957**, *49*, 1377–1383.
15. Ralph J, Lundquist K, Brunow G, Lu F, Kim H, Schatz PF, Marita JM, Hatfield RD, Ralph SA, Christensen JH, Boerjan W., *Phytochem Rev* **2004**, *3*, 29–60.
16. Isikgor FH, Becer CR., *Polymer Chemistry* **2015**, *6*, 4497-4459
17. Azadi P, Inderwildi OR, Farnood R, King DA., *Renew Sustain Energy Rev* **2013**, *21*, 506–523.
18. Faix O., *Holzforschung* **1991**, *45*, 21–28.
19. Pepper JM, Baylis PET, Adler E.,A., *Can J Chem* **1959**, *37* , 1241-1248.
20. Lapierre C, Pollet B, MacKay JJ, Sederoff RR., *J Agric Food Chem* **2000**, *48*. 2326–2331.
21. Zhang A, Lu F, Sun RC, Ralph J., *J Agric Food Chem* **2010**, *58*, 3446–3450.
22. Freudenberg K., *Science* **1965**, *148*, 595–600.
23. Mansfield SD, Kim H, Lu F, Ralph J., *Nat Protoc* **2012**, *7*, 1579–1589.
24. Karhunen P, Rummakko P, Sipilä J, Brunow G, Kilpeläinen I., *Tetrahedron Lett* **1995**, *36*, 4501–4504.
25. Chen Y, Chai L, Tang C, Yang Z, Zheng Y, Shi Y, Zhang H., *Bioresour Technol* **2012**, *123*, 682–685.
26. Del Río JC, Rencoret J, Prinsen P, Martínez ÁT, Ralph J, Gutiérrez A., *J Agric Food Chem* **2012**, *60*, 5922–5935.
27. Gall DL, Kontur WS, Lan W, Kim H, Li Y, Ralph J, Donohue TG, Noguera DR., *Appl Environ Microbiol* **2018**, *84*, 1-17.
28. Ogo Y, Ozawa K, Ishimaru T, Murayama T, Takaiwa F., *Plant Biotechnol J* **2013**, *11*, 734–746.
29. Chen H., *Lignocellulose Biorefinery Engineering* **2015**, 37–86.
30. Sluiter A, Hames B, Ruiz R, Scarlata C, Sluiter J, Templeton D, Crocker D., *Natl Renew Energy Lab* 2012, *17*.
31. Bunzel M, Schüßler A, Tchetseubu Saha G., *J Agric Food Chem* **2011**, *59*, 12506–12513.
32. Hatfield R, Fukushima RS., *Crop Sci* **2005**, *45*, 832–839.
33. Brauns FE., *J Am Chem Soc* **1939**, *61*, 2120–2127.

34. Marshall HB, Brauns F, Hibbert H., *Can J Res* **1935**, *13*, 103–113.
35. Björkman A., *Ind Eng Chem* **1957**, *49*, 1395–1398.
36. Kirk TK, Obst JR., *Methods Enzymol* **1988**, *161*, 87–101.
37. Kousini L, Fang Y, Paleologou M, Ahvazi B, Hawari J, Zhang Y, Wang XM., *Cellulose Chem Technol* **2011**, *45*, 515-520.
38. Zhu JY, Pan XJ, Wang GS, Gleisner R., *Bioresour Technol* **2009**, *100*, 2411–2418.
39. Lora JH, Glasser WG., *J Polym Environ* **2002**, *10*, 39–48.
40. Gordobil O, Moriana R, Zhang L, Labidi J, Sevastyanova O., *Ind Crops Prod* **2016**, *83*, 155–165.
41. Oliveira F, Ramires EC, Frollini E, Belgacem MN., *Ind Crops Prod* **2015**, *72*, 77–86.
42. Faruk O, Sain M, Farnood R, Pan Y, Xiao H., *J Polym Environ* **2014**, *22*, 279–288.
43. Gierer J, Wännström S., *Holzforschung* **1986**, *40*, 347–352.
44. Gierer J, Lindeberg O., *Acta Chemica Scandinavica* **1979**, *33*, 580–622.
45. Chakar FS, Ragauskas AJ., *Ind Crops Prod* **2004**, *20*, 131–141.
46. Samuel R, Pu Y, Raman B, Ragauskas AJ. *Appl Biochem Biotechnol* **2010**, *162*, 62–74.
47. Liitiä TM, Maunu SL, Hortling B, Toikka M, Kilpeläinen I., *J Agric Food Chem* **2003**, *51*, 2136–2143.
48. Zhu JY, Chandra MS, Gu F, Gleisner R, Reiner R, Sessions J, Marrs G, Gao J, Anderson D., *Bioresour Technol* **2015**, *179*, 390–397.
49. Pan X, Arato C, Gilkes N, Gregg D, Mabee W, Pye K, Xiao Z, Zhang X, Saddler J., *Biotechnol Bioeng* **2005**, *90*, 473–481.
50. Zhang L, Yan L, Wang Z, Laskar DD, Swita MS, Cort JR, Yang B., *Biotechnol Biofuels* **2015**, *8*, 1–14.
51. Alvarez-Vasco C, Ma R, Quintero M, Guo M, Geleynse S, Ramasamy KK, Wolcott M, Zhang X., *Green Chem* 2016, ;*18*, 5133–5141.
52. González M, Tejado Á, Peña C, Labidi J., *Ind Eng Chem Res* **2008**, *47*, 1903–1909.
53. Grande PM, Viell J, Theyssen N, Marquardt W, Domínguez De María P., Leitner W., *Green Chem* **2015**, *17*, 3533–3539.
54. vom Stein T, Grande PM, Kayser H, Sibilla F, Leitner W, Domínguez de María P., *Green Chem* **2011**, *13*, 1772-1777.
55. Kayser H, Rodríguez-Ropero F, Leitner W, Fioroni M, De María PD., *RSC Adv* **2013**, *3*, 9273–9278.
56. Chio C, Sain M, Qin W., *Renew Sustain Energy Rev* **2019**, *107*, 232–249.
57. Wong DWS., *Appl Biochem Biotechnol* **2009**, *157*, 174–209.
58. Pollegioni L, Tonin F, Rosini E., *FEBS J* **2015**, *282*, 1190–1213.
59. Giardina P, Faraco V, Pezzella C, Piscitelli A, Vanhulle S, Sannia G., *Cell Mol Life Sci* **2010**, *67*, 369–385.
60. Doyle WA, Blodig W, Veitch NC, Piontek K, Smith AT., *Biochemistry* **1998**, *37*, 15097–15105.
61. Kawai S, Nakagawa M, Ohashi H., *Enzyme Microb Technol* **2002**, *30*, 482–489.
62. Barreca AM, Fabbrini M, Galli C, Gentili P, Ljunggren S., *J Mol Catal B Enzym* **2003**, *26*, 105–110.
63. Li K, Xu F, Eriksson KEL., *Appl Environ Microbiol* **1999**, *65*, 2654–2660.
64. Solomon EI, Augustine AJ, Yoon J., *Dalt Trans* **2008**, *30*, 3921–3932.
65. Shraddha, Shekher R, Sehgal S, Kamthania M, Kumar A., *Enzyme Res* **2011**, *2011*, 1-11.
66. Dwivedi UN, Singh P, Pandey VP, Kumar A., *J Mol Catal B Enzym* **2011**, *68*, 117–128.
67. Arregui L, Ayala M, Gómez-Gil X, Gutiérrez-Soto G, Hernández-Luna CE, Herrera De Los Santos M, Levin L, Rojo-Dominguez A, Romero-Martinez D, Saparrat MCN,Trujillo-Roldan MA Valdez-Cruz NA., *Microb Cell Fact* **2019**, *18*, 1–33.

68. Reiss R, Ihssen J, Richter M, Eichhorn E, Schilling B, Thöny-Meyer L., *PLoS One* **2013**, *8*, 1-10).
69. Rodríguez Couto S, Toca Herrera JL., *Biotechnol Adv* **2006**, *24*, 500–513.
70. Paliwal R, Rawat AP, Rawat M, Rai JPN., *Appl Biochem Biotechnol* **2012**, *167*, 1865–1889.
71. Martínez ÁT, Speranza M, Ruiz-Dueñas FJ, Ferreira P, Camarero S, Guillén F, Martinez MJ, Gutierrez A, del Rio J., *Int Microbiol* **2005**, *8*, 195–204.
72. Wang W, Wen X., *J Environ Sci* **2009**, *21*, 218–222.
73. Nie G, Reading NS, Aust SD., *Biochem Biophys Res Commun* **1998**, *249*, 146–150.
74. Haemmerli SD, Leisola MSA, Fiechter A., *FEMS Microbiol Lett* **1986**, *35*, 33–36.
75. Glenn JK, Gold MH., *Arch Biochem Biophys* **1985**, *242*, 329–341.
76. Kumar V, Chandra R., *World J Microbiol Biotechnol* **2018**, *34*, 1–18.
77. Hammel KE, Cullen D., *Curr Opin Plant Biol* **2008**, *11*, 349–355.
78. Hofrichter M., *Enzyme Microb Technol* **2002**, 30, 454–466.
79. Hofrichter M, Ullrich R, Pecyna MJ, Liers C, Lundell T., *Appl Microbiol Biotechnol* **2010**, *87*, 871–897.
80. Reading NS, Aust SD., *Biotechnol Prog* **2000**, *16*, 326–333.
81. Miyazaki C, Takahashi H., *FEBS Lett* **2001**, *509*, 111–114.
82. Ufot UF, Akpanabiatu MI., *Am J Mol Biol* **2012**, *2*, 359–370.
83. Sridhar M., *J Dairy Vet Anim Res* **2016**, *4*, 277–280.
84. Ruiz-Dueñas FJ, Morales M, Pérez-Boada M, Choinowski T, Martínez MJ, Piontek K, Martinez AT., *Biochemistry* **2007**, *46*, 66–77.
85. Pérez-Boada M, Ruiz-Dueñas FJ, Pogni R, Basosi R, Choinowski T, Martínez MJ, Piontek K, Martinez AT., *J Mol Biol* **2005**, *354*, 385–402.
86. Roberts JN, Singh R, Grigg JC, Murphy MEP, Bugg TDH, Eltis LD., *Biochemistry* **2011**, *50*, 5108–5119.
87. Van Bloois E, Pazmiño DET, Winter RT, Fraaije MW*., Appl Microbiol Biotechnol* **2010**, *86*, 1419–1430.
88. Ahmad M, Taylor CR, Pink D, Burton K, Eastwood D, Bending GD, Bugg TDH., *Mol Biosyst* **2010**, *6*, 815–821.
89. Colpa DI, Fraaije MW, Van Bloois E., *J Ind Microbiol Biotechnol* **2014**, *41*, 1–7.
90. Yoshida T, Tsuge H, Konno H, Hisabori T, Sugano Y., *FEBS J* **2011**, *278*, 2387–2394.
91. Linde D, Ruiz-Dueñas FJ, Fernández-Fueyo E, Guallar V, Hammel KE, Pogni R, Martinez AT., *Arch Biochem Biophys* **2015**, *574*, 66–74.
92. Silveira CM, Moe E, Fraaije M, Martins LO, Todorovic S., *RSC Adv* **2020**, *10*, 11095–11104.
93. Lončar N, Drašković N, Božić N, Romero E, Simić S, Opsenica I, Vujcic Z, Fraaije MW., *Catalysts* **2019**, *9*, 1–10.
94. Atkinson HJ, Babbitt PC., *Biochemistry* **2009**, *48*, 11108–11116.
95. Masai E, Nishikawa S, Morohoshi N, Haraguchi T., *FEBS Lett* **1989**, *249*, 348-352.
96. Picart P, De María PD, Schallmey A., *Front Microbiol* **2015**, *6*, 1–8.
97. Masai E, Katayama Y, Kubota S, Kawai S, Yamasaki M, Morohoshi N., *FEBS Lett* **1993**, *323*, 135–140.
98. Masai E, Katayama Y, Nishikawa S, Fukuda M., *J Ind Microbiol Biotechnol* **1999**, *23*, 364–373.
99. Masai E, Katayama Y, Fukuda M., *Biosci Biotechnol Biochem* **2007**, *71*, 1–15.
100. Tanamura K, Abe T, Kamimura N, Kasai D, Hishiyama S, Otsuka Y, Nakamura M, Kajita S, Katayama Y, Fukuda M, Masai E., *Biosci Biotechnol Biochem* **2011**, *75*, 2404–2407.

101. Masai E, Kubota S, Katayama Y, Kawai S, Yamasaki M, Morohoshi N., *Biosci Biotechnol Biochem* **1993**, *57*, 1655–16549.
102. Sato Y, Moriuchi H, Hishiyama S, Otsuka Y, Oshima K, Kasai D,Nakamura M, Ohana S, Katayama Y, Fukuda M, Masai E*., Appl Environ Microbiol* **2009**, *75*, 5195–5201.
103. Masai E, Ichimura A, Sato Y, Miyauchi K, Katayama Y, Fukuda M., *J Bacteriol* **2003**, *185*, 1768–1775.
104. Gall DL, Kim H, Lu F, Donohue TJ, Noguera DR, Ralph J., *J Biol Chem* **2014**, *289*, 8656–8667.
105. Picart P, Sevenich M, Domínguez de María P, Schallmey A., *Green Chem* **2015**, *17*, 4931–4940.
106. Otsuka Y, Sonoki T, Ikeda S, Kajita S, Nakamura M, Katayama Y., *Eur J Biochem* **2003**, *270*, 2353–2362.
107. Gall DL, Ralph J, Donohue TJ, Noguera DR., *Environ Sci Technol* **2014**, *48*, 12454–12463.
108. Ohta Y, Nishi S, Hasegawa R, Hatada Y., *Sci Rep* **2015**, *5*, 1-14.
109. Kontur WS, Bingman CA, Olmsted CN, Wassarman DR, Ulbrich A, Gall DL, Smith RW, Yusko LM, Fox BG, Noguera DR, Coon JJ, Donohue TJ., *J Biol Chem* **2018**, *293*, :4955–4968.
110. Picart P, Müller C, Mottweiler J, Wiermans L, Bolm C, Dominguez de Maria P, Schallmey A., *ChemSusChem* **2014**, *7*, 3164–3171.
111. Voß H, Heck CA, Schallmey M, Schallmey A*., Appl Environ Microbiol* **2020**, *86*, 1-15.
112. Helmich KE, Pereira JH, Gall DL, Heins RA, McAndrew RP, Bingman C, Deng K, Holland KC, Noguera DR, Simmons BA, Sale KL, Ralph J, Donohue TJ, Adams PD, Phillips GN., *J Biol Chem* **2016**, *291*, 5234–5246.
113. Husarcikova J, Schallmey A., *J Biotechnol* **2019**, *293*, 1-7.
114. Allocati N, Federici L, Masulli M, Di Ilio C., *FEBS J* **2009**, *276*, 58–75.
115. Kontur WS, Olmsted CN, Yusko LM, Niles A V., Walters KA, Beebe ET, Vander Meulen KA, Karlen SD, Gall DL, Noguera DR, Donohue TJ., *J Biol Chem* **2019**, *294*, 1877–1890.
116. Prates ET, Crowley MF, Skaf MS, Beckham GT., *J Phys Chem B* **2019**, *123*, 10142–10151.
117. Meux E, Prosper P, Masai E, Mulliert G, Dumarçay S, Morel M, Didierjean C, Gelhaye E, Favier F., *FEBS Lett* **2012**, *586*, 3944–3950.
118. Pereira JH, Heins RA, Gall DL, McAndrew RP, Deng K, Holland KC, Donohue TJ, Noguera DR, Simmons BA, Sale KL, Ralph J, Adams PD., *J Biol Chem* **2016**, *291*, 10228–10238.
119. Reiter J, Strittmatter H, Wiemann LO, Schieder D, Sieber V., *Green Chem* **2013**, *15*, 1373-1381.
120. Ohta Y, Hasegawa R, Kurosawa K, Maeda AH, Koizumi T, Nishimura H, Okada H, Qu Ch, Saito K, Watanabe T, Hatada Y., *ChemSusChem* **2017**, *10*, 425–433.
121. Crestini C, Melone F, Sette M, Saladino R., *Biomacromolecules* **2011**, *12*, :3928–3935.
122. Picart P, Liu H, Grande PM, Anders N, Zhu L, Klankermayer J, Leitner W, Dominguez de Maria P, Schwanenberg U, Schallmey A., *Appl Microbiol Biotechnol* **2017**, 1–11.
123. Chen K, Arnold FH., *Biotechnology* **1991**, *9*, 1073–1077.
124. Pritchard L, Corne D, Kell D, Rowland J, Winson M., *J Theor Biol* **2005**, *234*, 497–509.
125. Martínez R, Schwaneberg U., *Biol Res* **2013**, *46*, 395–405.
126. Bornscheuer UT, Pohl M., *Curr Opin Chem Biol* **2001**, *5*, 137–143.
127. Reetz MT, Carballeira JD., *Nat Protoc* **2007**, *2*, 891–903.
128. Reetz MT, Kahakeaw D, Lohmer R., *ChemBioChem* **2008**, *9*, 1797–804.
129. Schulz F, Leca F, Hollmann F, Reetz MT., *Beilstein J Org Chem* **2005**, *1*, 1–9.
130. Gall DL, Ralph J, Donohue TJ, Noguera DR., *Curr Opin Biotechnol* **2017**, *45*, 120–126.
131. Chen C, Fujimoto Y, Girdaukas G, Sih CJ., *J Am Chem Soc* **1982**, *104*, 7294-7299

132. Schallmey A, Den Besten G, Teune IGP, Kembaren RF, Janssen DB. *Appl Microbiol Biotechnol* **2011**, *89*, 1475–1485.
133. Gustafson C, Tagesson C., *Brit J Ind Med* **1985**, *42*, 591–595.
134. Kruger NJ, Von Schaewen A., *Curr Opin Plant Biol* **2003**, *6*, 236–246.
135. Smirnova G, Muzyka N, Oktyabrsky O., *Microbiol Res* **2012**, *167*, 166–172.
136. Fahey RC, Brown WC, Adams WB, Worsham MB., *J Bacteriol* **1978**, *133*, 1126–1129.
137. Goldstein DB., *Ann Emerg Med* **1986**, *15*, 1013–1018.
138. Ganske F, Bornscheuer UT., *Biotechnol Lett* **2006**, *28*, 465–469.
139. Henderson GB, Murgolo NJ, Kuriyan J, Osapay K, Kominos D, Berry A, Scrutton NS, Hinchliffe NW, Perham RN, Cerami A., *Proc Natl Acad Sci USA* **1991**, *88*, 8769–8773.
140. Trott O, Olson AJ., *J Comput Chem* **2009**, *31*, 455-461.
141. Acids A, Dougherty DA., *J Nutr* **2007**, *137*, 1504–1508.
142. Casalone E, Allocati N, Ceccarelli I, Masulli M, Rossjohn J, Parker MW, Carmine DI., *FEBS Lett* **1998**, *423*, 122–124.
143. Federici L, Masulli M, Bonivento D, Di Matteo A, Gianni S, Favaloro B, Di Ilio C, Allocat N., *Biochem J* **2007**, *403*, 267–274.
144. Nishida M, Harada S, Noguchi S, Satow Y, Inoue H, Takahashi K., *J Mol Biol* **1998**, *281*, 135–147.
145. Teale FW, Weber G., *Biochem J* **1957**, *65*, 476–482.
146. Ghisaidoobe ABT, Chung SJ., *Int J Mol Sci* **2014**, *15*, 22518–22538.
147. Matveeva E, Morisseau C, Goodrow M, Mullin C, Hammock B., *Curr Pharm Biotechnol* **2009**, *10*, 589–599.
148. Akbar SMD, Sreeramulu K, Sharma HC., *J Bioenerg Biomembr* **2016**, *48*, 241–247.
149. Jernström B, Funk M, Frank H, Mannervik B, Seidel A., *Carcinogenesis* **1996**, *17*, 1491–1498.
150. Ivarsson Y, Mannervik B., *Protein Eng Des Sel* **2005**, *18*, 607–616.
151. Hu X, Ji X, Srivastava SK, Xia H, Awasthi S, Nanduri B, Awasthi YC, Zimniak P, Singh SV., *Arch Biochem Biophys* **1997**, *45*, 32–38.
152. Brunmark A, Cadenas E., *Free Radic Biol Med* **1989**, *6*, 149–165.
153. Sørensen HP, Mortensen KK., *J Biotechnol* **2005**, *115*, 113–128.
154. Yegane Sarkandy S, Shojaosadati SA, Khalilzadeh R, Sadeghizadeh M., Iran *J Chem Chem Eng* **2013**, *32*, 127–131.
155. Wang H, Wang F, Wang W, Yao X, Wei D, Cheng H, Deng Z., *PLoS One* **2014**, *9*, 1–11.
156. Rosano GL, Ceccarelli EA., *Front Microbiol* **2014**, *5*, 1–17.
157. Christensen DG, Orr JS, Rao C V, Wolfe J., *Appl Environ Microbiol* **2017**, *83*, 1–13.
158. Akerboom TPM, Sies H., *Methods Enzymol* **1981**, *77*, 373–382.
159. Studier FW, Moffatt BA., *J Mol Biol* **1986**, *189*, 113–130.
160. Dumon-Seignovert L, Cariot G, Vuillard L., *Protein Expr Purif* **2004**, *37*, 203–206.
161. Miroux B, Walker JE., *J Mol Biol* **1996**, *260*, 289–298.
162. Phue JN, Sang JL, Trinh L, Shiloach J., *Biotechnol Bioeng* **2008**, *101*, 831–836.
163. Haeyoung Jeong, Hyun Ju Kim SJL*., Genome Announc* **2015**, *3*, 1.
164. Wagner S, Klepsch MM, Schlegel S, Appel A, Draheim R, Tarry M, Hogbom M, van Wijk KJ, Slotbloom DJ, Persson JO, de Gier JW., *Proc Natl Acad Sci USA* **2008**, *105*, 14371–14376.
165. Kwon SK, Kim SK, Lee DH, Kim JF., *Sci Rep* **2015**, *5*, 1-13.
166. Wagner S, Baarst L, Ytterberg AJ, Klussmerer A, Wagner CS, Nord O, Nyrgen PA, van Wijk KJ, de Gier JW., *Mol Cell Proteomics* **2007**, *6*, 1527–1550.

167. Gubellini F, Verdon G, Karpowich NK, Luff JD, Boël G, Gauthier N, Handelman SK, Ades SE, Hunt JF., *Mol Cell Proteomics* **2011**, *10*, 1–17.
168. Slonczewski JL, Rosent BP, Alger JR, Macnab RM., *Proc Natl Acad Sci USA* **1981**, *78*, 6271–6275.
169. Rubio E, Fernandez-Mayorales A, Klibanov AM*., J Am Chem Soc* **1991**, *113*, 695–696.
170. Bridiau N, Issaoui N, Maugard T., *Biotechnol Prog* **2010**, *26*, 1278–1289.
171. Weber FJ, De Bont JAM., *Biochim Biophys Acta* **1996**, *1286*, 225–245.
172. Jain MK, Wu NM., *J Membr Biol* **1977**, *34*, 157–201.
173. Meyer F, Keller P, Hartl J, Gröninger OG, Kiefer P, Vorholt JA., *Nat Commun* **2018**, *9*, 1-10.
174. Müller JEN, Meyer F, Litsanov B, Kiefer P, Potthoff E, Heux S, Quax WJ, Wendisch VF, Brautaset T, Portais JCh, Vorholt JA., *Metab Eng* **2015**, *28*, 190–201.
175. Bennett RK, Gonzalez JE, Whitaker WB, Antoniewicz MR, Papoutsakis ET. *Metab Eng* **2018**, *45*, 75–85.
176. Whitaker WB, Jones JA, Bennett RK, Gonzalez JE, Vernacchio VR, Collins SM, Palmer MA, Schmidt S, Antoniewicz MR, Koffas MA, Papoutsakis ET., *Metab Eng* **2017**, *39*, 49–59.
177. Gonzalez JE, Bennett RK, Papoutsakis ET, Antoniewicz MR., *Metab Eng* **2018**, *45*, 67–74.
178. Dyrda G, Boniewska-Bernacka E, Man D, Barchiewicz K, Słota R., *Mol Biol Rep* **2019**, *46*, 3225–3232.
179. Hernández Fernández FJ, Pérez de los Ríos A, Quesada-Medina J, Sánchez-Segado S*., ChemBioEng Rev* **2015**, *2*, 44–53.
180. Carrea G., *Trends Biotechnol* **1984**, *2*, 102–106.
181. Eş I, Vieira JDG, Amaral AC., *Appl Microbiol Biotechnol* **2015**, *99*, 2065–2082.
182. Markošová K, Husarčíková J, Halásová M, Kourist R, Rosenberg M, Stloukal R, Zajoncova L, Rebros M., *Catalysts* **2018**, *8*, 1-11.
183. Greer S, Perham RN., *Biochemistry* **1986**, *25*, 2736–2742.
184. Ricklefs E, Girhard M, Urlacher VB., *Microb Cell Fact* **2016**, *15*, 1–11.
185. Lin B, Tao Y., *Microb Cell Fact* **2017**, *16*, 1–12.
186. Friedman M, Henika PR, Mandrell RE., *J Food Prot* **2002**, *65*, 1545–1560.
187. Cole SPC, Deeley RG., *Trends Pharmacol Sci* **2006**, *27*, 438–446.
188. Sharma R, Awasthi S, Zimniak P, Awasthi YC., *Acta Biochim Pol* **2000**, *47*, 751–762.
189. König J, Nies AT, Cui Y, Leier I, Keppler D., *Biochim Biophys Acta* **1999**, *1461*, 377–394.
190. Jeon H, Yoon S, Ahsan MM, Sung S, Kim GH, Sundaramoorthy U, Rhee SK, Yun H., *Catalysts* **2017**, *7*, 1-13.
191. Pittman MS, Robinson HC, Poole RK., *J Biol Chem* **2005**, *280*, 32254–32261.
192. Suzuki H, Koyanagi T, Izuka S, Onishi A, Kumagai H., *J Bacteriol* **2005**, *187*, 5861–5867.
193. Xiao Z, Lv C, Gao C, Qin J, Ma C, Liu Z, Liu P, Li L, Xu P., *PLoS One* **2010**, *5*, 1–6.
194. Lengauer T, Rarey M., *Curr Opin Struct Biol* **1996**, *6*, 402–406.
195. Meng, X. Y., Zhang, H. X., Mezei, M., & Cui M., *Curr Comput Aided Drug Des* **2011**, *7*, 146–157.
196. Kitchen DB, Decornez H, Furr JR, Bajorath J., *Nat Rev Drug Discov* **2004**, *3*, 935–949.
197. Wei BQ, Weaver LH, Ferrari AM, Matthews BW, Shoichet BK., *J Mol Biol* **2004**, *337*, 1161–1182.
198. Huang SY, Zou X., *Int J Mol Sci* **2010**, *11*, 3016–3034.
199. Cross JB, Thompson DC, Rai BK, Baber JC, Fan KY, Hu Y, Humblet Ch., *J Chem Inf Model* **2009**, *49*, 1455–1474.
200. Elokely KM, Doerksen RJ., *J Chem Inf Model* **2013**, *53*, 1934–1945.
201. Tuccinardi T, Poli G, Romboli V, Giordano A, Martinelli A., *J Chem Inf Model* **2014**, 54, 2980–2986.

202. Wang G, Zhu W., *Future Med Chem* **2016**, *8*, 1-4.
203. Cole JC, Murray CW, Nissink JWM, Taylor RD, Taylor R., *Proteins Struct Funct Genet* **2005**, *60*, 325–332.
204. Tao X, Huang Y, Wang C, Chen F, Yang L, Ling L, Che Z., *Int J Food Sci Technol* **2020**, *55*, 33–45.
205. Wrenbeck EE, Azouz LR, Whitehead TA., *Nat Commun* **2017**, *8*, 1–10.
206. Oakley A., *Drug Metab Rev* **2011**, *43*, 138–351.
207. Busk PK, Lange L., *Appl Environ Microbiol* **2013**, *79*, 3380–3391.
208. Lee J, Goodey NM., *Chem Rev* **2011**, *111*, 7595–7624.
209. Yang G, Hong N, Baier F, Jackson CJ, Tokuriki N., *Biochemistry* **2016**, *55*, 4583–4593.
210. Oue S, Okamoto A, Yano T, Kagamiyama H., *J Biol Chem* **1999**, *274*, 2344–2349.
211. Schmidt M, Hasenpusch D, Kähler M, Kirchner U, Wiggenhorn K, Langel W, Bornscheuer UT., *ChemBioChem* **2006**, *7*, 805–809.
212. Omari K El, Liekens S, Bird LE, Balzarini J, Stammers DK., *Mol Pharmacol* **2006**, *69*, 1891–1896.
213. Tomatis PE, Rasia RM, Segovia L, Vila AJ., *Proc Natl Acad Sci USA* **2005**, *102*, 13761–13766.
214. Tyukhtenko S, Rajarshi G, Karageorgos I, Zvonok N, Gallagher ES, Huang H, Vemuri K, Hudgens JW, Ma X, Nasr ML, Pavlopoulos S, Makriyannis A., *Sci Rep* **2018**, *8*, 1–17.
215. Wang J, Yan K, Wang W, Zhou Y., *Bioresour Bioprocess* **2020**, *7*, 1-8.
216. Deponte M., *Biochim Biophys Acta* **2013**, *1830*, 3217–3266.
217. Armstrong RN., *Chem Res Toxicol* **1997**, *10*, 2–18.
218. Bartlett GJ, Porter CT, Borkakoti N, Thornton JM., *J Mol Biol* **2002**, *324*, 105–121.
219. Ribeiro AJM, Holliday GL, Furnham N, Tyzack JD, Ferris K, Thornton JM., *Nucleic Acids Res* **2018**, *46*, 618–623.
220. Perez-Rizquez C, Abian O, Palomo JM., *Chem Commun* **2019**, *55*, 12928–12931.
221. Miyanokoshi M, Yokosawa T, Wakasugi K., *J Biol Chem* **2018**, *293*, 8428–8438.
222. Palamuru S, Dellas N, Pearce SL, Warden AC, Oakeshott JG, Pandey G., *Appl Environ Microbiol* **2015**, *81*, 8164–8176.
223. Acerenza L, Mizraji E., *Biochim Biophys Acta* **1997**, *1339*, 155–166.
224. Goldbeter A., *Biophys Chem* **1976**, *4*, 159–169.
225. Abeliovich H., *Biophys J* **2005**, *89*, 76–79.
226. Jakobson I, Warholm M, Mannervik B*., J Biol Chem* **1979**, *254*, 7085–7089.
227. Danielson UH, Mannervik B., *Biochem J* **1985**, *231*, 263–267.
228. Tahir MK, Mannervik B., *J Biol Chem* **1986**, *261*, 1048–1051.
229. Labrou NE, Mello L V., Clonis YD., *Eur J Biochem* **2001**, *268*, 3950–3957.
230. Ricci G, Lo Bello M, Caccuri AM, Pastore A, Nuccetelli M, Parker MW, Federici G., *J Biol Chem* **1995**, *270*, 1243-1248.
231. Bogaards JJP, Venekamp JC, Van Bladeren PJ., *Chem Res Toxicol* **1997**, *10*, 310–317.
232. Kong KH, Inoue H, Takahashi K., *Protein Eng Des Sel* **1993**, *6*, 93–99.
233. Caccuri AM, Antonini G, Ascenzi P, Nicotra M, Nuccetelli M, Mazzetti AP, Federici G, Lo Bello M, Ricci G., *J Biol Chem* **1999**, *274*, 19276–19280.
234. Quesada-Soriano I, Barón C, García-Maroto F, Aguilera AM, García-Fuentes L., *Biochemistry* **2013**, *52*, 1980–1989.
235. Valenzuela-Chavira I, Contreteas-Vergara CA, Arvizu-Flores AA, Serrano-Posada H, Lopez-Savara AA, Garcia-Orozco KD, Hernandez-Paredes J, Rudino-Pinera E, Stojanoff V, Sotelo-Mundo RR, Islas-Osuna MA., *Biochimie* **2018**, *135*, 35–45.

236. Neuefeind T, Huber R, Dasenbrock H, Prade L, Bieseler B., *J Mol Biol* **1997**, *274*, 446–453.
237. Pivnenko K, Pedersen GA, Eriksson E, Astrup TF., *Waste Manag* **2015**, *44*, 39–47.
238. Giulivo M, Lopez de Alda M, Capri E, Barceló D., *Environ Res* **2016**, *151*, 251–264.
239. Oehlmann J, Schulte-Oehlmann U, Kloas W, Jagnytsch O, Lutz I, Kusk KO, Wollenberger L, Santos EM, Paull GC, van Look KJW, Tyler ChR., *Philos Trans R Soc B Biol Sci* **2009**, *364*, 2047–2062.
240. Tharp AP, Maffini M V., Hunt PA, VandeVoort CA, Sonnenschein C, Soto AM., *Proc Natl Acad Sci USA* **2012**, *109*, 8190–8195.
241. Christiansen S, Axelstad M, Boberg J, Vinggaard AM, Pedersen GA, Hass U., *Reproduction* **2014**, *147*, 477–487.
242. Inadera H., *Int J Med Sci* **2015**, *12*, 926–936.
243. Brede C, Fjeldal P, Skjevrak I, Herikstad H., *Food Addit Contam* **2003**, *20*, 684–689.
244. Cao XL, Corriveau J., *J Agric Food Chem* **2008**, *56*, 6378–6381.
245. Huang YQ, Wong CKC, Zheng JS, Bouwman H, Barra R, Wahlström B, Neretin L, Wong MH., *Environ Int* **2012**, *42*, 91–99.
246. Diepens M, Gijsman P., *Polym Degrad Stab* **2007**, *92*, 397–406.
247. Hoekstra EJ, Simoneau C., *Crit Rev Food Sci Nutr* **2013**, *53*, 386–402.
248. Carwile JL, Luu HT, Bassett LS, Driscoll DA, Yuan C, Chang JY, Ye X, Calafat AM, Michels KB., *Environ Health Perspect* **2009**, *117*, 1368–1372.
249. Guerra P, Kim M, Teslic S, Alaee M, Smyth SA., *J Environ Manage* **2015**, *152*, 192–200.
250. Gassara F, Brar SK, Verma M, Tyagi RD., *Chemosphere* **2013**, *92*, 1356–1360.
251. Huang Q, Weber WJ., *Environ Sci Technol* **2005**, *39*, 6029–6036.
252. Sakurai A, Toyoda S, Sakakibara M., *Biotechnol Lett* **2001**, *23*, 995–998.
253. Yamanaka H, Moriyoshi K, Ohmoto T, Ohe T, Sakai K., *J Biosci Bioeng* **2008**, *105*, 157–160.
254. Alomar MS, Boyles DA., *ACS Omega* **2019**, *4*, 22363–22372.
255. Sasaki M, Maki JI, Oshiman KI, Matsumura Y, Tsuchido T., *Biodegradation* **2005**, *16*, 449–459.
256. Oshiman KI, Tsutsumi Y, Nishida T, Matsumura Y., *Biodegradation* **2007**, *18*, 247–255.
257. Inoue D, Hara S, Kashihara M, Murai Y, Danzl E, Sei K, Tsunoi S, Fujita M, Ike M., *Appl Environ Microbiol* **2008**, *74*, 352–358.
258. Zhou NA, Kjeldal H, Gough HL, Nielsen JL., *Environ Sci Technol* **2015**, *49*, 12231–12241.
259. Jia Y, Eltoukhy A, Wang J, Li X, Hlaing TS, Aung MM, Nwe MT, Lamraoui I, Yan Y., *Int J Mol Sci* **2020**, *21*, 1-15.
260. Arslan ÖÇ, Parlak H, Boyacioglu M, Karaslan MA, Katalay S., *Fresenius Environ Bull* **2014**, *23*, 2525–2530.
261. Sravani J, Padmaja K, Prasad PE, Kumari BP., *Int J Meat Sci* **2015**, *6*, 1–5.
262. Ficociello G, Gerardi V, Uccelletti D, Setini A., *Environ Sci Pollut Res* **2020**, 1-13
263. Liu Q, Wang J, Lin B, Cheng ZY, Bai M, Shi S, Huang XX, Song SJ., *Bioorg Chem* **2019**, *84*, 269–275.
264. Buckler JN, Banwell MG, Kordbacheh F, Parish CR, Santiago FS, Khachigian LM., *ACS Omega* **2017**, *2*, 7375–7388.
265. Liu X, Fu J, Yao XJ, Yang J, Liu L, Xie TG, Jiang PCh, Jiang ZH, Zhu GY., *J Nat Prod* **2018**, *81*, 1333–1342.
266. Yoganantharajah P, Ray AP, Eyckens DJ, Henderson LC, Gibert Y., *BMC Biotechnol* **2018**, *18*, 1–7.
267. https://patents.google.com/patent/US1856567A/en

268. Kovach ME, Phillips RW, Elzer PH, Roop RM, Peterson KM., *Biotechniques* **1994**, *16*, 800-802.
269. Voß H., *Identification, characterization and engineering of glutathione-dependent, lignin-degrading enzymes* (doctoral thesis), **2019**, TU Braunschweig, Germany.
270. Terbrack J., *Untersuchung der Glutathion-abhängigen β-Etherasen LigE und LigF durch Protein Engineering* (bachelor thesis), **2018**, TU Braunschweig, Germany.
271. https://chisure.en.made-in-china.com/product/KSkmyEQBnWpw/China-L-Glutathione-Reduced-High-Quality-Best-Price-CAS-70-18-8.html
272. http://www.itwreagents.com/united-states/en/product/l-glutathione-reduced-ibiochemicai/A2084
273. Peters T., *Untersuchung Glutathion-abhängiger Enzyme zur Spaltung von β-O-4 Aryletherbindungen in Ligninmodellsubstraten* (bachelor thesis), **2019**, TU Braunschweig, Germany.
274. https://sites.google.com/site/miller00828/in/solvent-polarity-table
275. Nelson DL, Cox MM., *Lehninger Principles of Biochemistry* (*Seventh Edition*) **2017**, ISBN-10: 1-4641-2611-9.
276. https://eur-lex.europa.eu/eli/reg/2018/213/oj#document1

7 Appendix

7.1 List of figures

7.2 List of tables

7.3 List of abbreviation

AA	Amino acid
ABTS	2,2'-azino-bis-(3-ethylbenzothiazoline-6-sulfonic acid)
ACN	Acetonitrile
Amp	Ampicillin
APS	Ammonium persulfate
ATP	Adenosine triphosphate
AV	Acetoveratrone
BLAST	Basic Local Alignment Searching Tool
bp	Base pairs
B-PER	Bacterial Protein Extraction Reagent
CFE	Cell free extract
CV	Column volume
DCM	Dichloromethane
DFT	Density Functional Theory
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DyP	Dye-decolorizing peroxidase
e.g.	*exempli gratia* (for example)
ESI-MS	Electro Spray Ionization Mass Spectrometry
et al.	*et alii (and others)*
EtOAc	Ethyl acetate
EtOH	Ethanol
FPLC	Fast protein liquid chromatography
Fwd	Forward
G	Guaiacyl unit
GSH	Glutathione

GSSG	Oxidized glutathione
GST	Glutathione transferase
H	*p*-hydroxyphenyl unit
HBT	1-hydroxybenzotriazole
HPLC	High performance chromatography
IMAC	Immobilized metal affinity chromatography
IPCC	Intergovernmental Panel on Climate Change
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IR	Infrared radiation
kan	Kanamycin
kb	Kilobase
k_{cat}	Turnover number
kDa	Kilodalton
Kpi	Potassium phosphate buffer
LB	Luria broth
LC-MS	Liquid chromatography-mass spectrometry
LED	Light emitting diode
LiP	Lignin peroxidase
M	Marker
MilliQ	Ultrapure water
MnP	Manganese peroxidase
MST	Microscale thermophoresis
NAD	Nicotinamide dinucleotide
NADP	Nicotinamide dinucleotide phosphate
NBS	*N*-bromsuccinimid
NTA	Nitrilotriacetic acid
NMR	Nuclear magnetic resonance
OD_{600}	Optical density at λ = 600 nm
PCR	Polymerase chain reaction

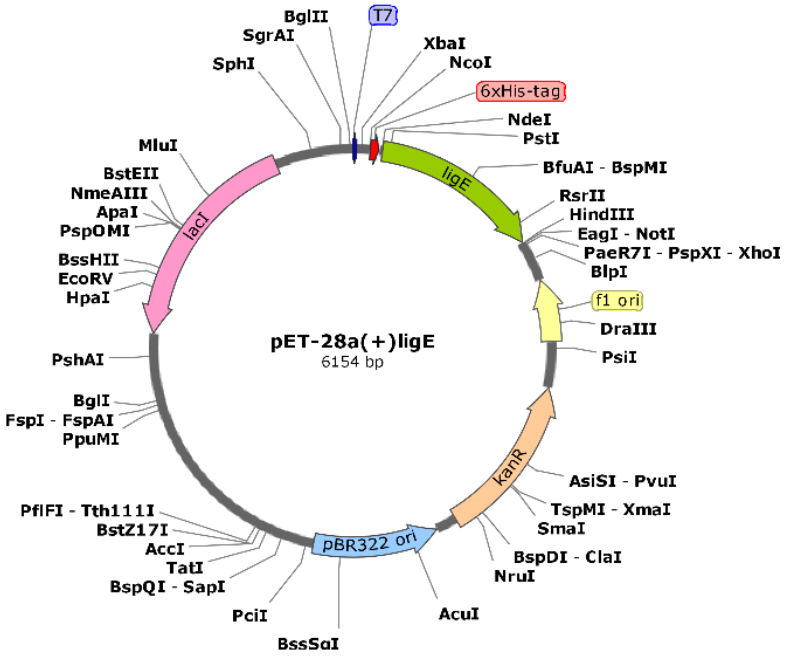
PDB	Protein data bank
PMSF	Phenylmethylsulfonyl fluoride
rac	Racemic
rpm	Revolutions per minute
rev	Reverse
S	Syringyl unit
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SOC	Super optimal broth with Catabolite repression
SSM	Site-saturation mutagenesis
TAE	Tris base, acetic acid and EDTA
TB	Terrific broth
TEMED	Tetramethylethylenediamine
TEMPO	2,2,6,6-tetramethylpiperidine-1-oxyl
TLC	Thin Layer Chromatography
T_m	Melting temperature
Tris	Tris(hydroxymethyl)aminomethane
UV	Ultraviolet
VA	Veratryl alcohol
VG	Veratryl glycerone
VP	Versatile peroxidase
WT	Wild type

7.3.1 Abbreviation of amino acids

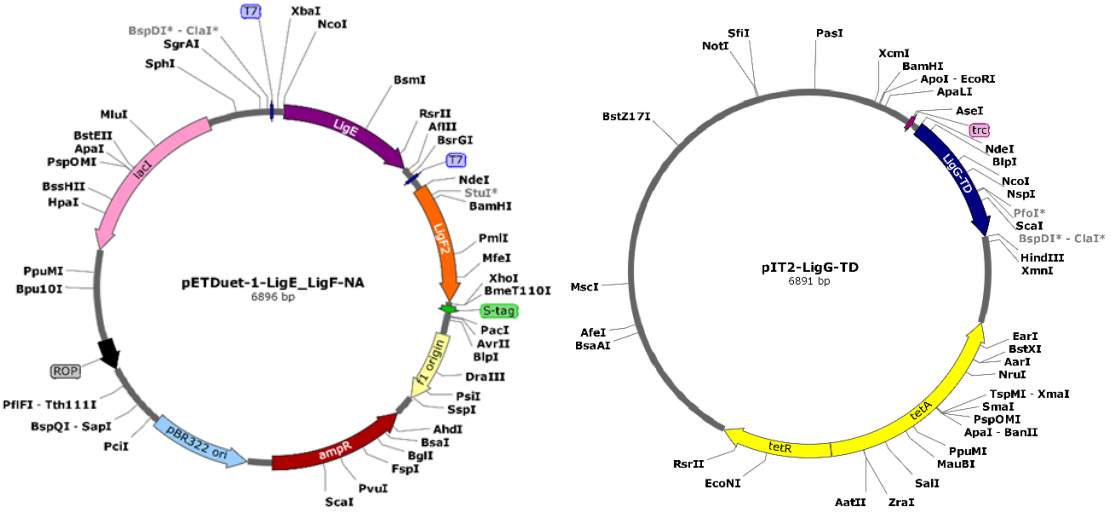
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C

Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

7.4 Vector maps



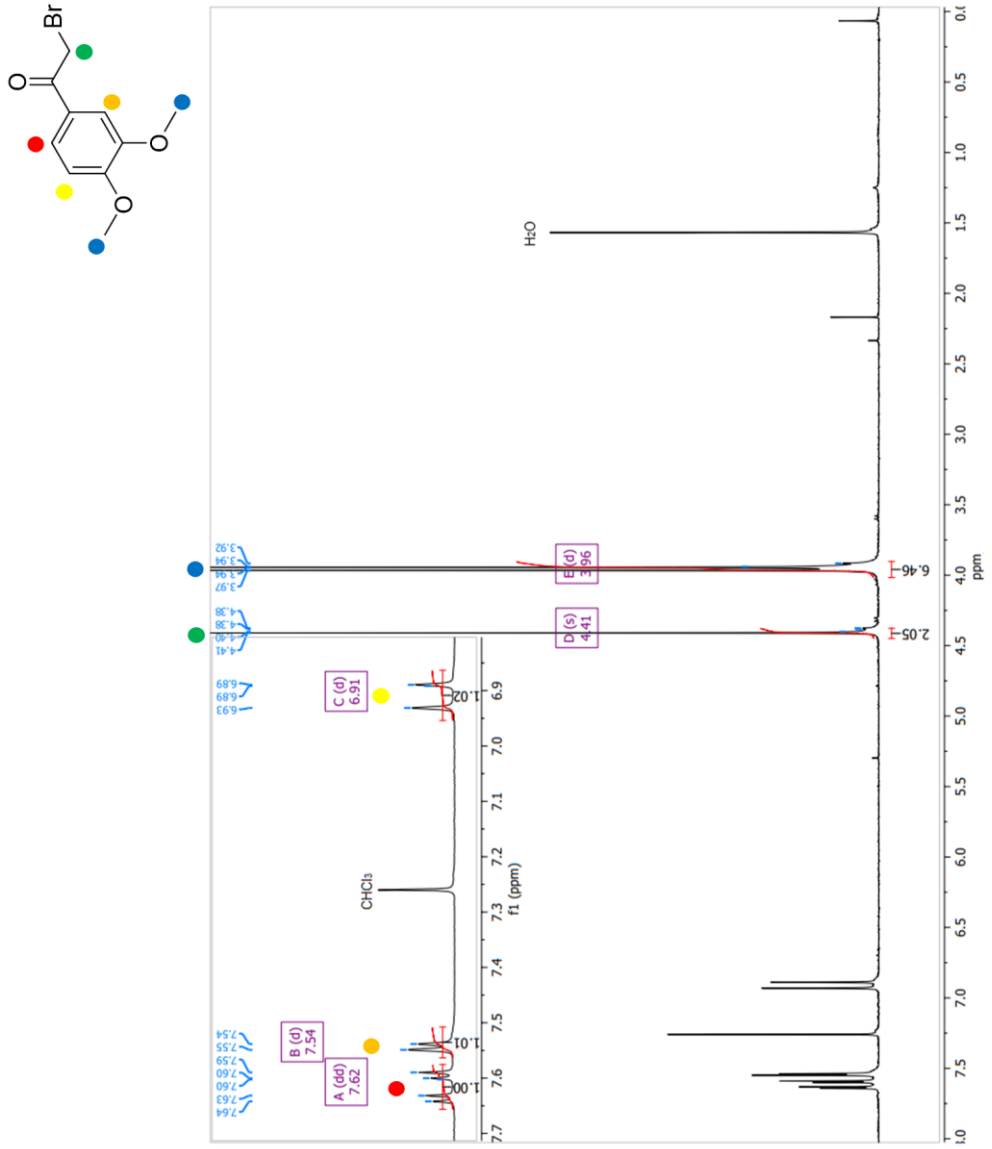
Supplementary Figure. 7.1: Gene *ligE* cloned into the pET28a(+) expression vector using restriction sites NdeI and HindIII.



Supplementary Figure 7.2: Vector maps of pETDuet1_*ligE_ligF* harbouring β-etherase genes *ligE* (MCS1) and *ligF-NA* (MCS2) as well as vector pIT2_*ligG-TD* carrying the gene encoding glutathione lyase LigG-TD.

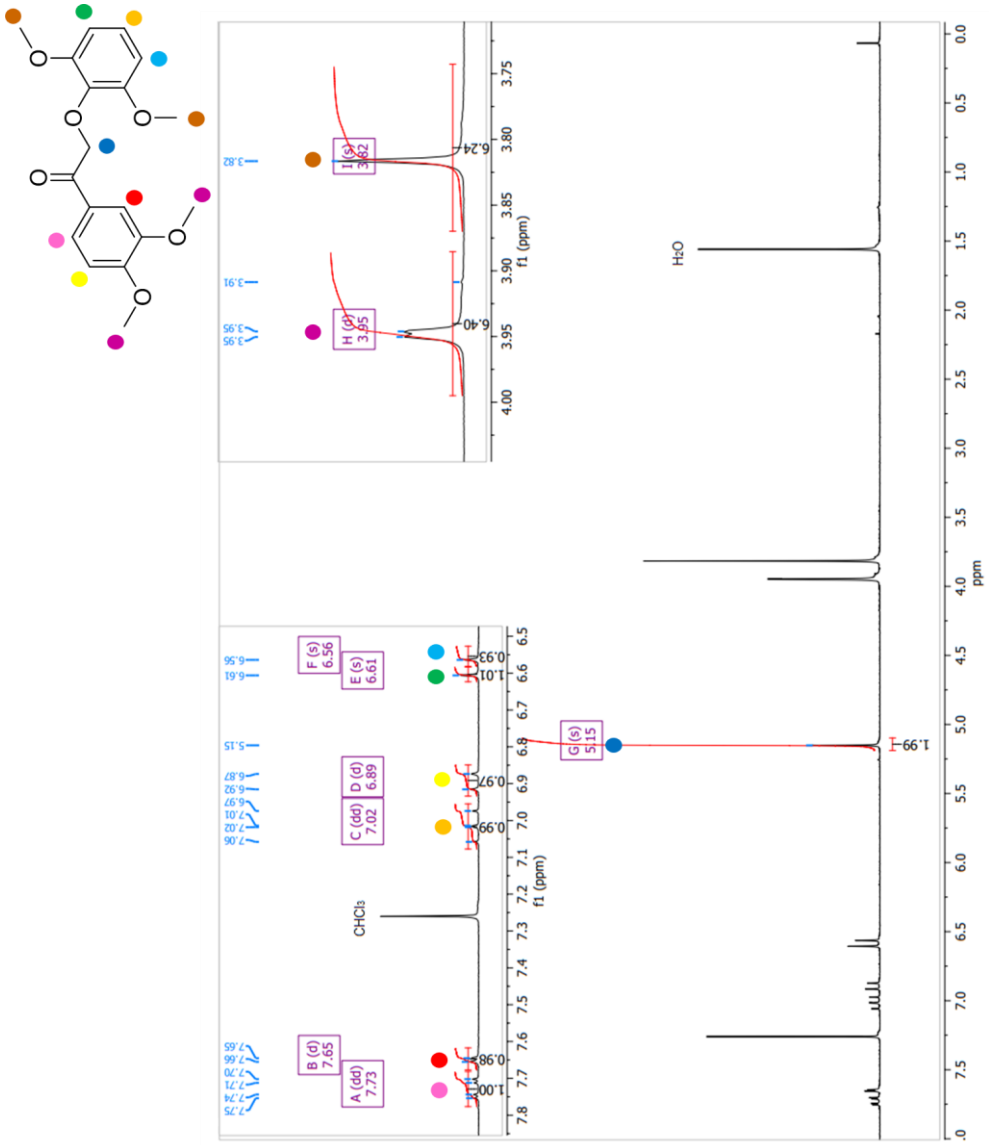
7.5 ^{1}H NMR Spectra

Compound B



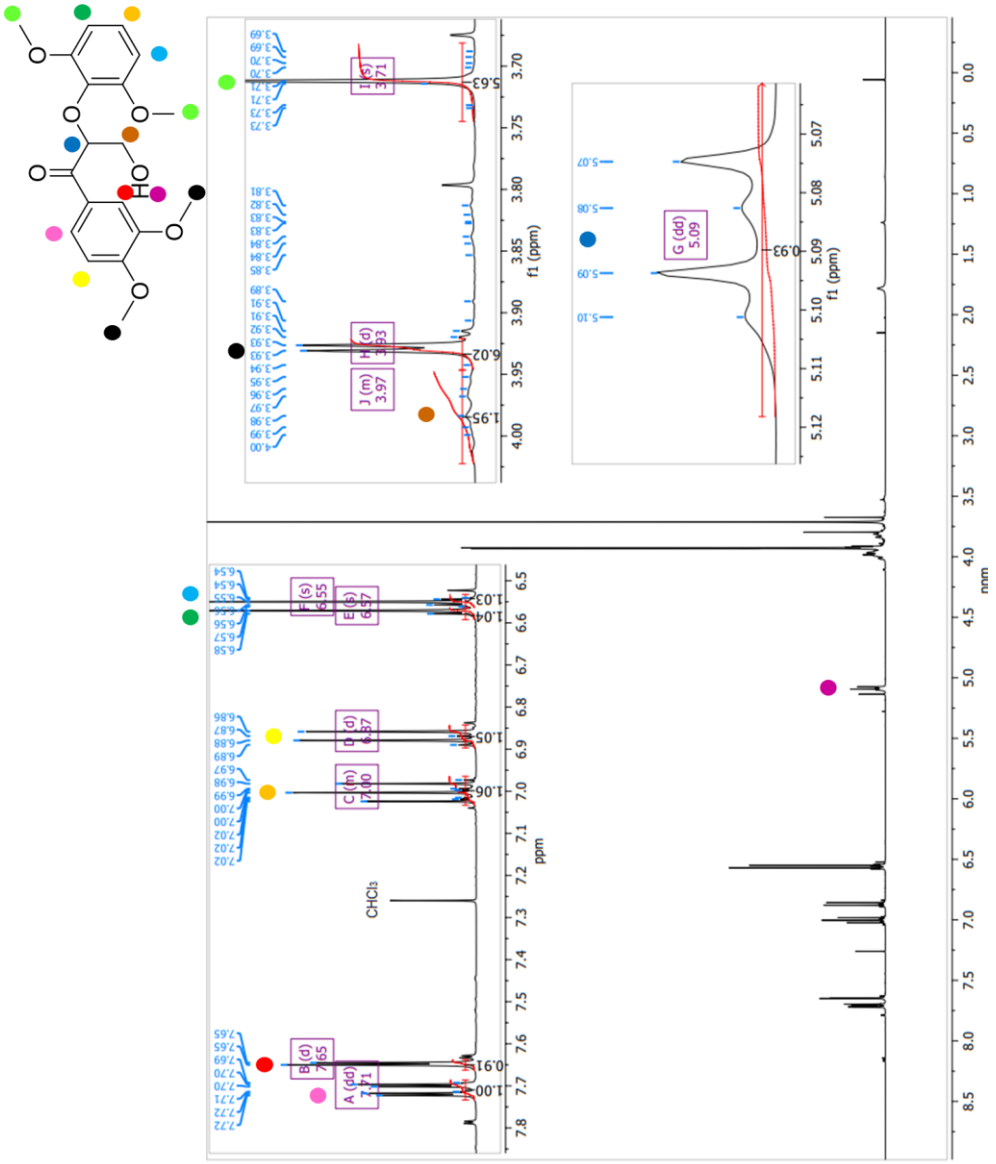
Supplementary Figure 7.3: ^{1}H NMR (300 MHz, $CDCl_3$) of 3,4-dimethoxyacetophenon (**B**) (Spectrum analysed with MestReNova Lite).

Compound 2



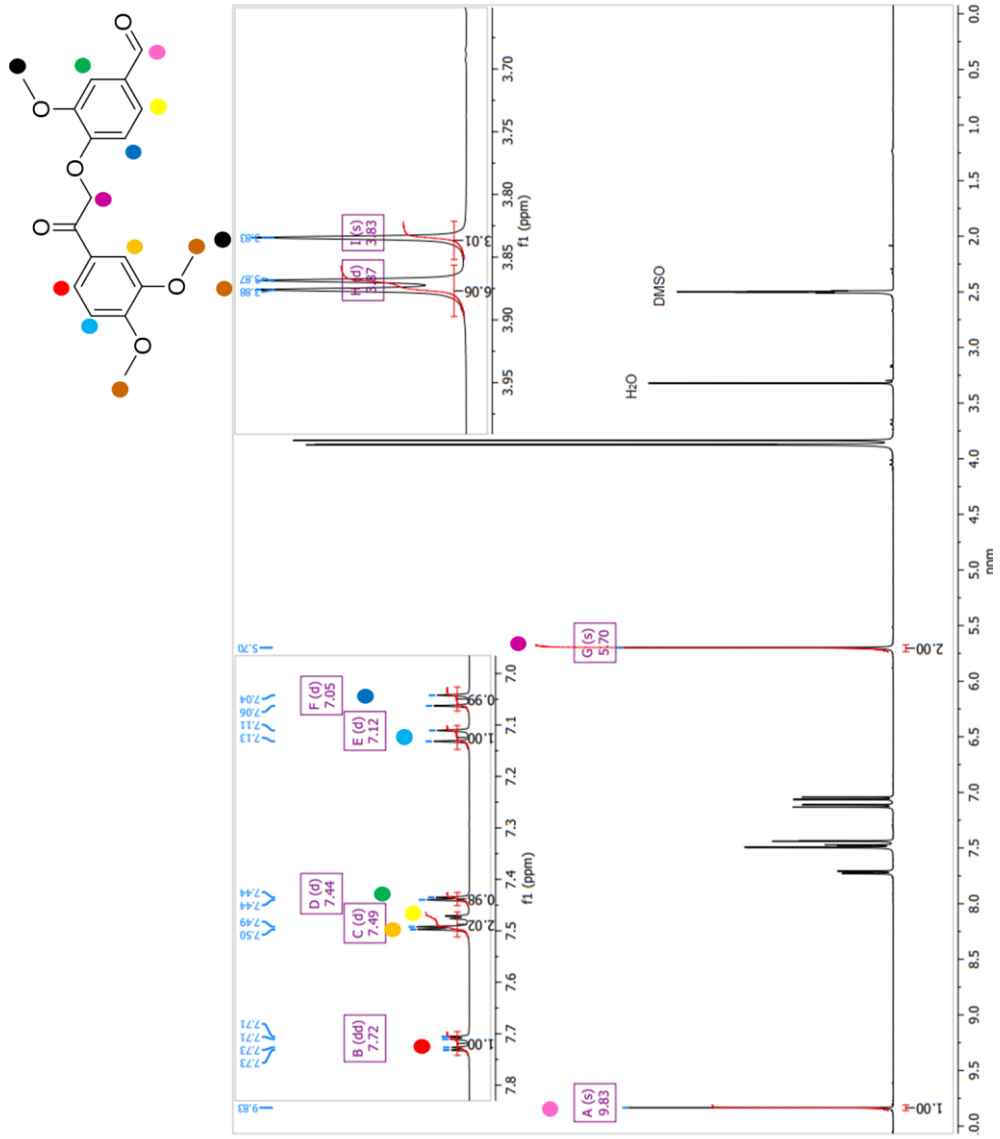
Supplementary Figure 7.4: ^{1}H NMR (300 MHz, $CDCl_3$) of β-(2,6-dimethoxyphenoxy)-α-(3,4-dimethoxyphenyl)-ethan-1-on (**2**) (Spectrum analysed with MestReNova Lite).

Compound 1



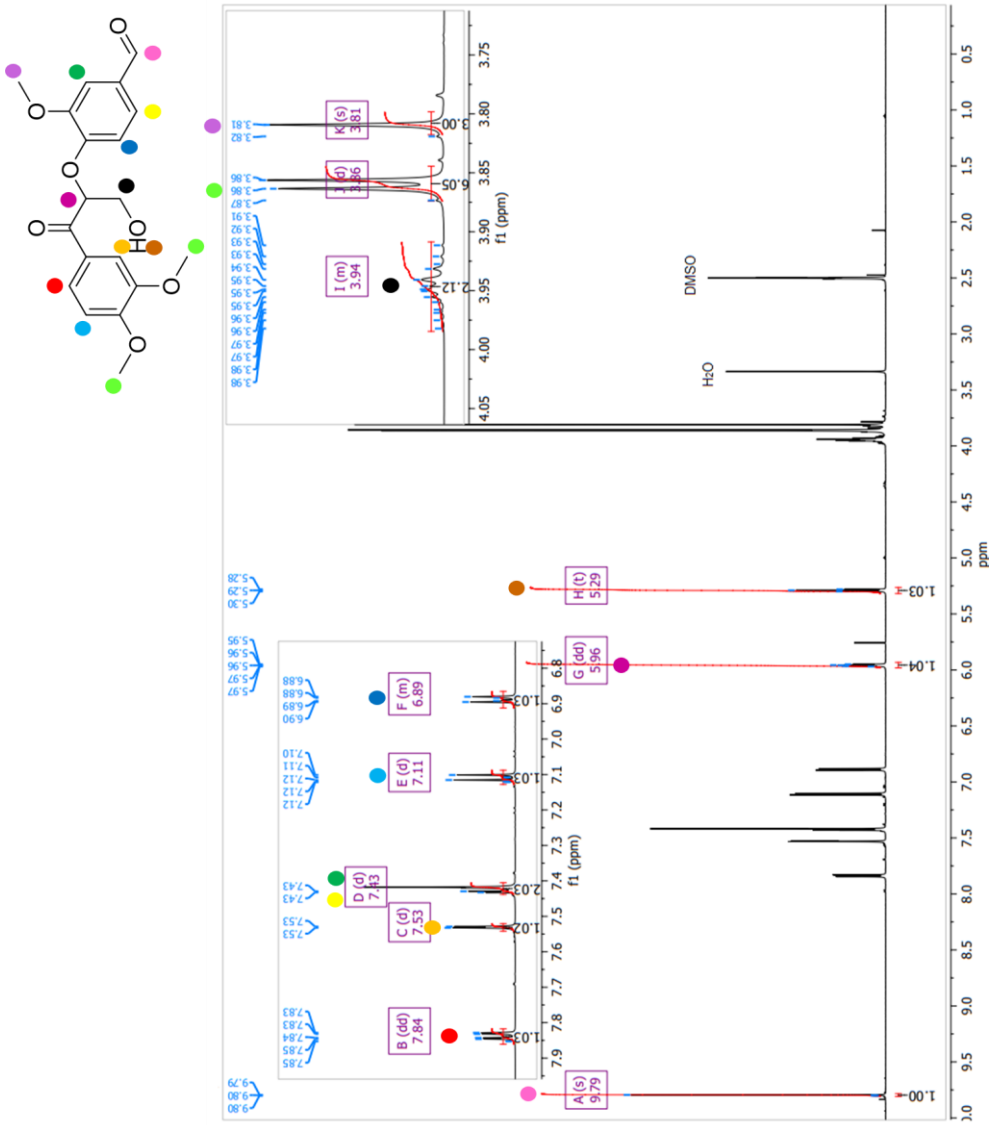
Supplementary Figure 7.5: ^{1}H NMR (300 MHz, $CDCl_3$) of β-(2,6-dimethoxyphenoxy)-α-veratrylglycerone (**1**) (Spectrum analysed with MestReNova Lite).

Compound C



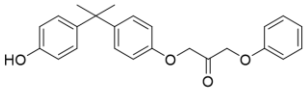
Supplementary Figure 7.6: ^{1}H NMR (300 MHz, dDMSO) of 4-(2-(3,4-dimethoxyphenyl)-2-oxoethoxy)-3-methoxybenzaldehyde (**C**) (Spectrum analysed with MestReNova Lite).

Compound 3

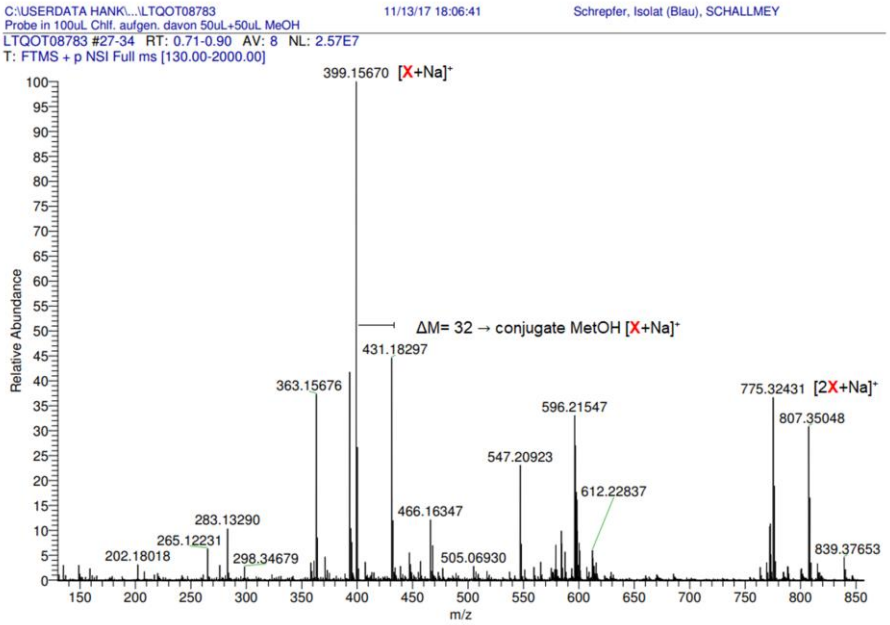


Supplementary Figure 7.7: ^{1}H NMR (300 MHz, dDMSO) of (β-vanillinyl-α-veratrylglycerone (**3**) (Spectrum analysed with MestReNova Lite).

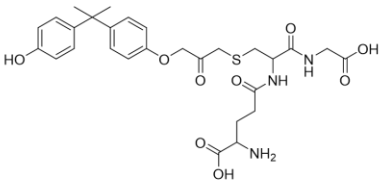
7.6 LC-MS, MS analysis



X = Molecular Weight: 376.45



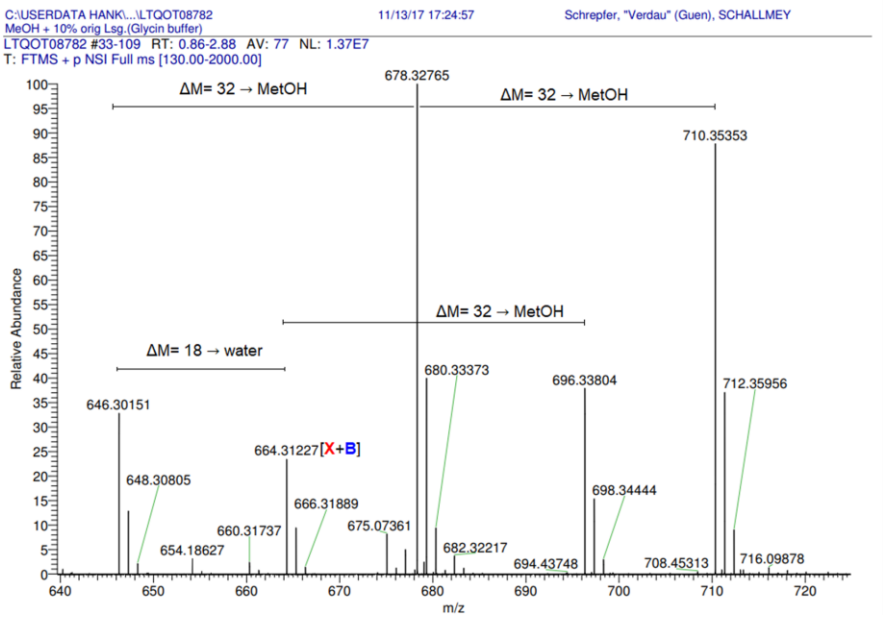
Supplementary Figure 7.8: MS spectra of the substrate in the reaction, where mixture **30** was tested.



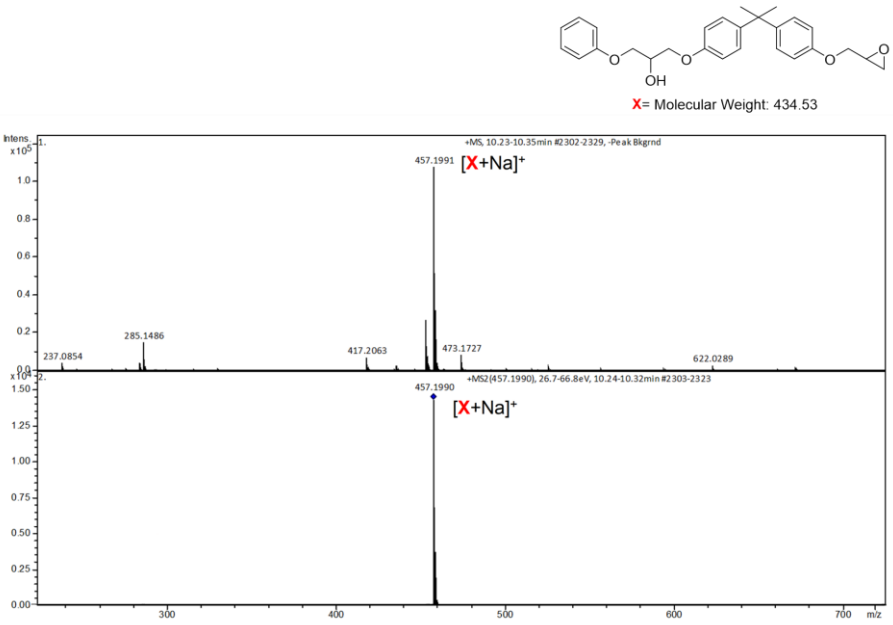
X= Molecular Weight: 589.66

Glycine buffer

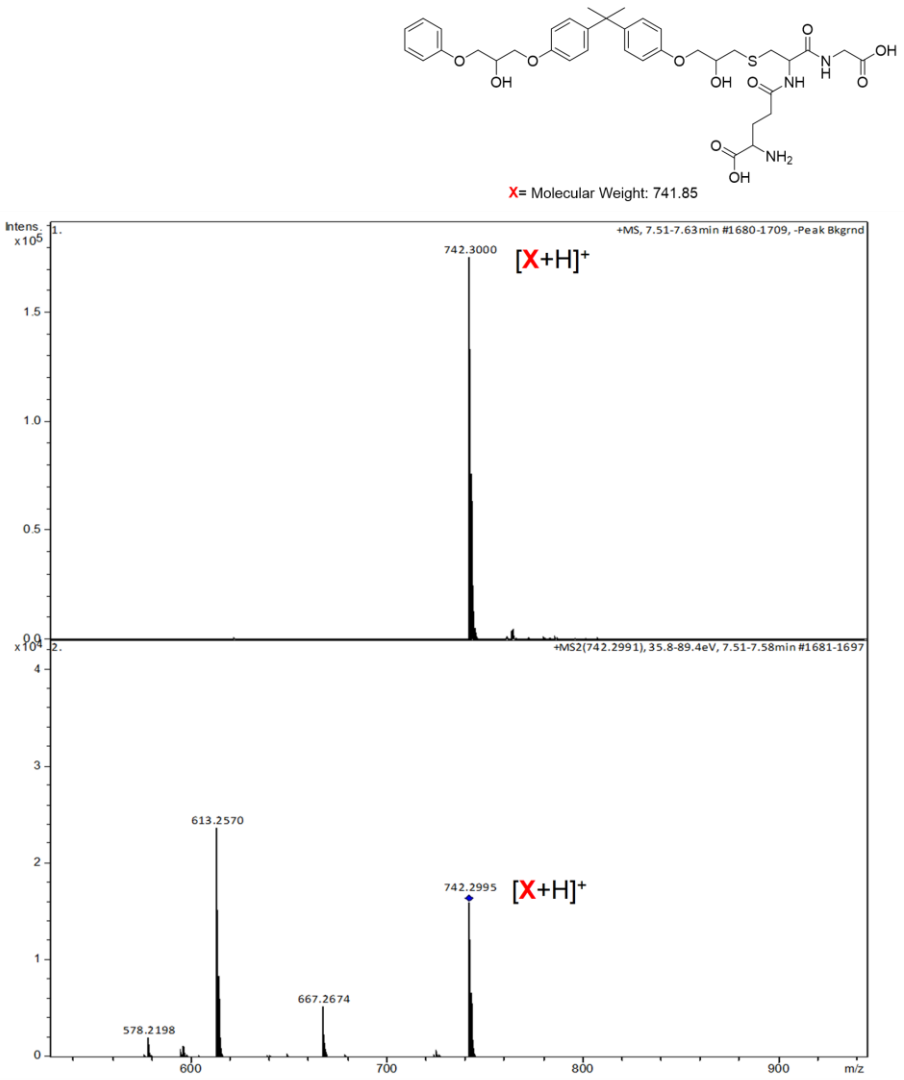
B= Molecular Weight: 75.07



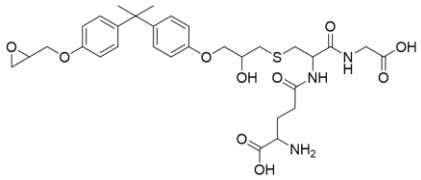
Supplementary Figure 7.9: MS spectra of the product in the reaction, where mixture **30** was tested.



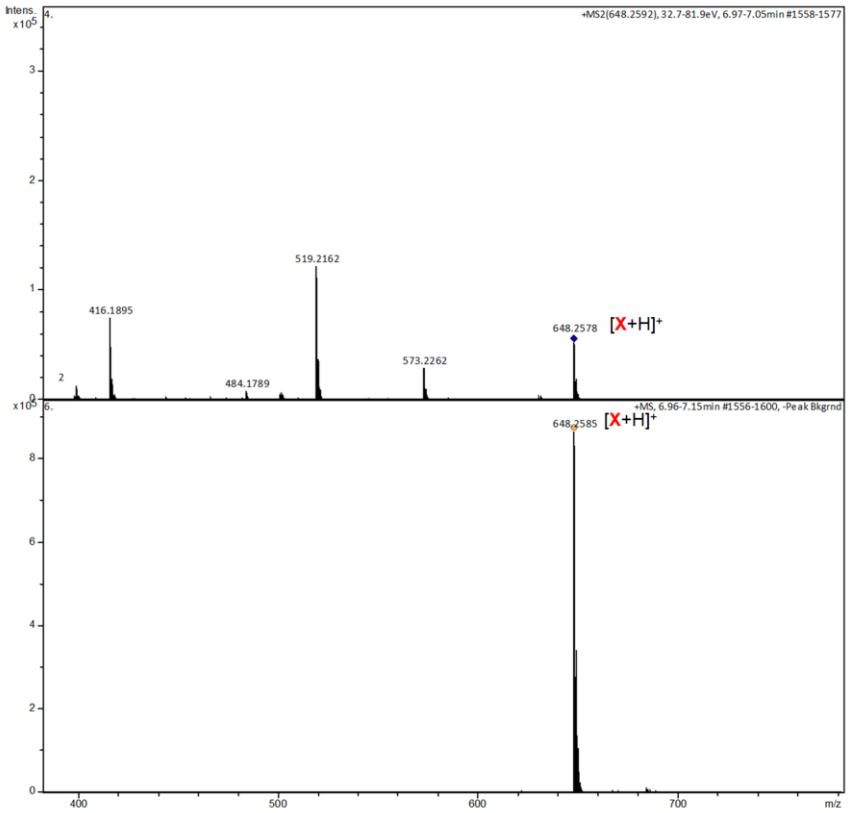
Supplementary Figure 7.10: LC-MS spectra of the substrate in the reaction, where mixture **31** was tested.



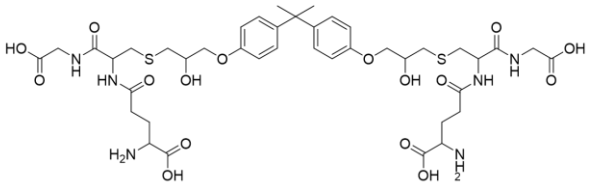
Supplementary Figure 7.11: LC-MS spectra of the product in the reaction, where mixture **31** was tested.



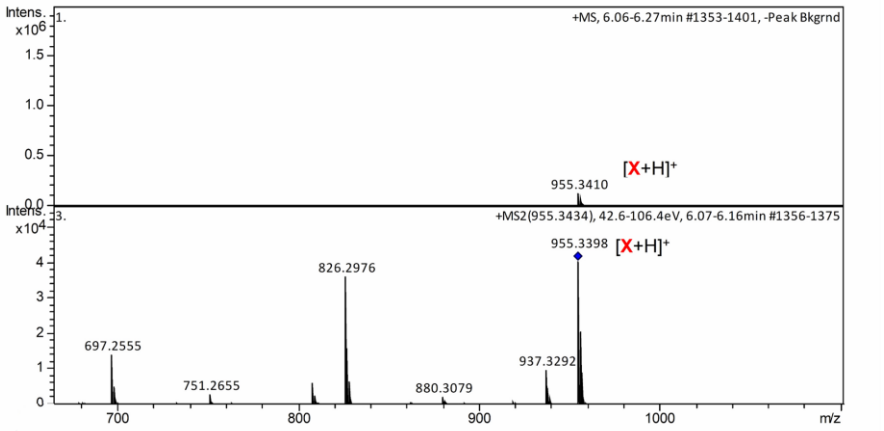
X= Molecular Weight: 647.74



Supplementary Figure 7.12: LC-MS spectra of the intermediate in the reaction, where mixture **32** was tested.

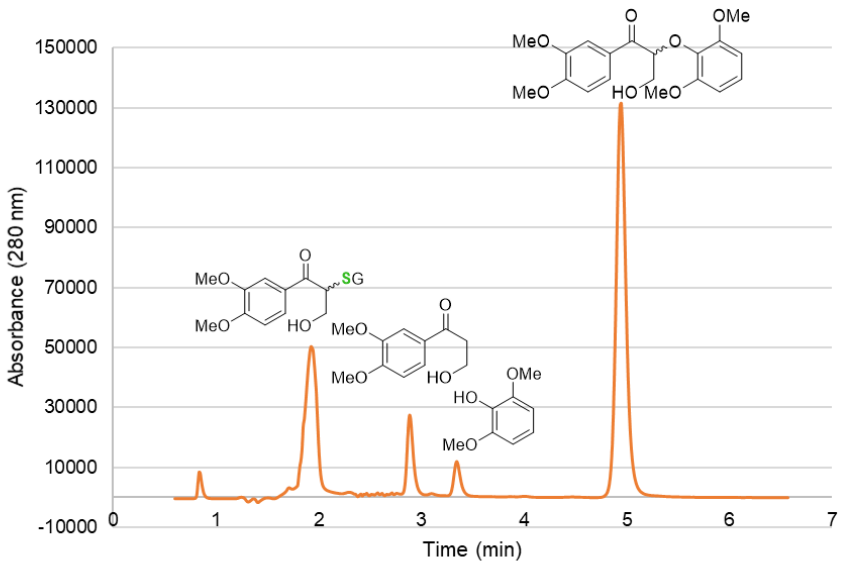


X= Molecular Weight: 954.34



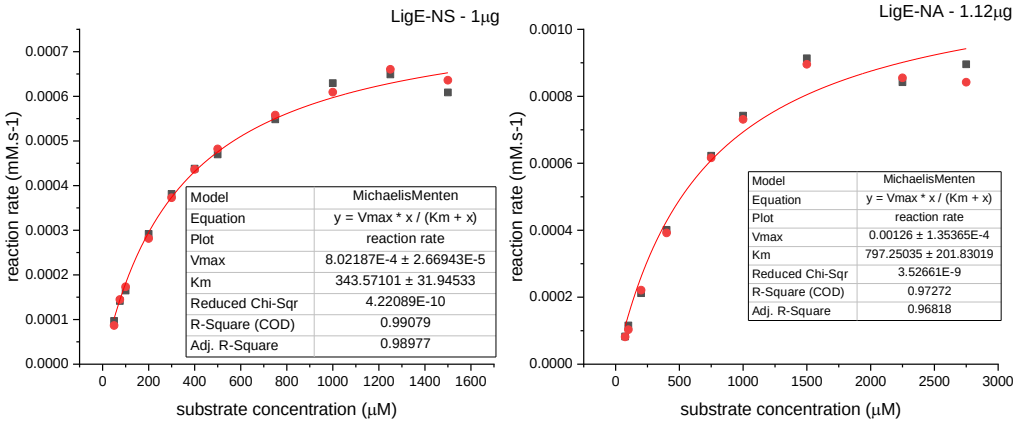
Supplementary Figure 7.13: LC-MS spectra of the final product in the reaction, where mixture **32** was tested.

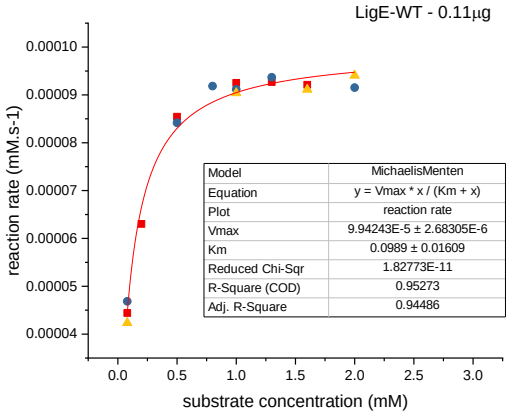
7.7 HPLC chromatogram

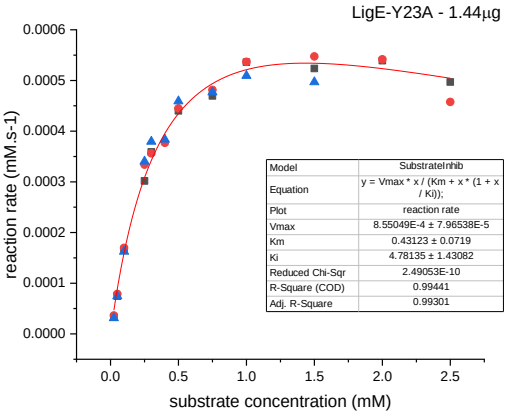


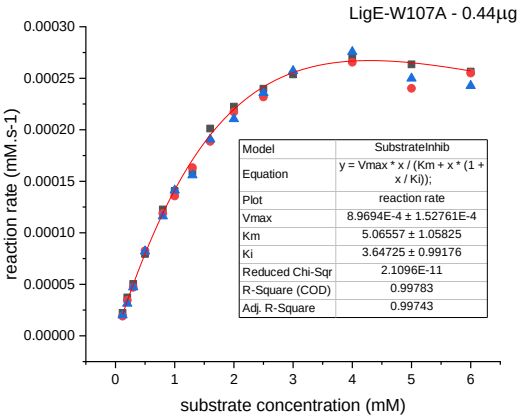
Supplementary Figure 7.14: HPLC chromatogram of achiral analysis of substrate **1** with the corresponding intermediates and final product when in reaction with β-etherases and glutathione lyases.

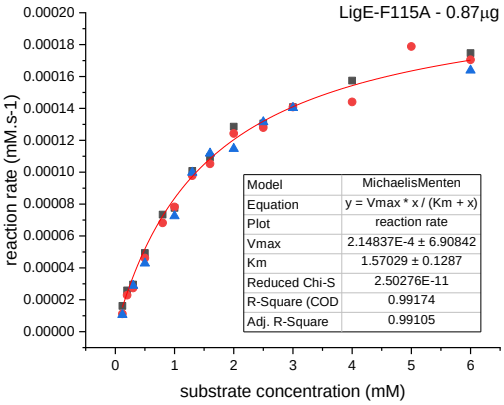
7.8 Kinetic measurements

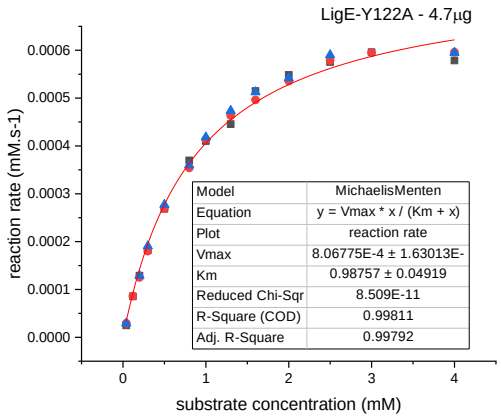


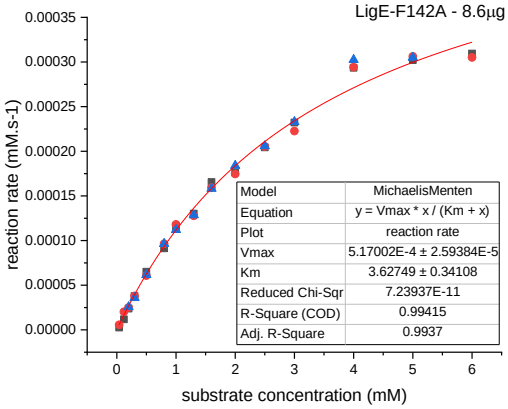

LigE-WT - 0.11µg
reaction rate (mM.s-1)
substrate concentration (mM)
Model MichaelisMenten
Equation y = Vmax * x / (Km + x)
Plot reaction rate
Vmax 9.94243E-5 ± 2.68305E-6
Km 0.0989 ± 0.01609
Reduced Chi-Sqr 1.82773E-11
R-Square (COD) 0.95273
Adj. R-Square 0.94486

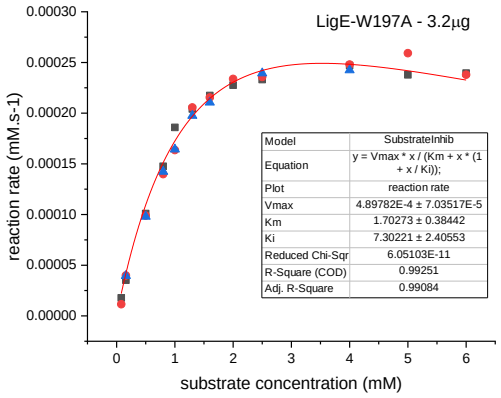

LigE-Y23A - 1.44µg
reaction rate (mM.s-1)
substrate concentration (mM)
Model SubstrateInhib
Equation y = Vmax * x / (Km + x * (1 + x / Ki));
Plot reaction rate
Vmax 8.55049E-4 ± 7.96538E-5
Km 0.43123 ± 0.0719
Ki 4.78135 ± 1.43082
Reduced Chi-Sqr 2.49053E-10
R-Square (COD) 0.99441
Adj. R-Square 0.99301

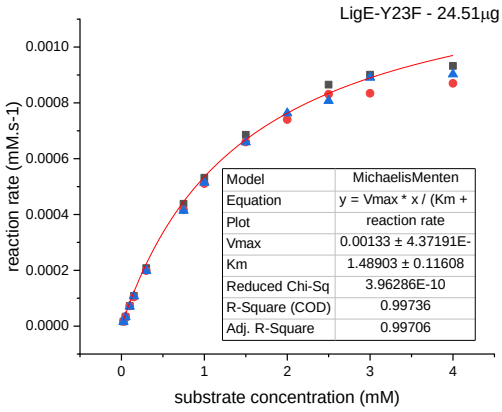

LigE-W107A - 0.44µg
reaction rate (mM.s-1)
substrate concentration (mM)
Model SubstrateInhib
Equation y = Vmax * x / (Km + x * (1 + x / Ki));
Plot reaction rate
Vmax 8.9694E-4 ± 1.52761E-4
Km 5.06557 ± 1.05825
Ki 3.64725 ± 0.99176
Reduced Chi-Sqr 2.1096E-11
R-Square (COD) 0.99783
Adj. R-Square 0.99743

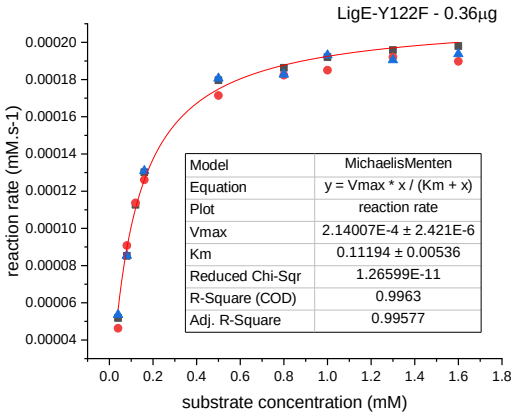

LigE-F115A - 0.87µg
reaction rate (mM.s-1)
substrate concentration (mM)
Model MichaelisMenten
Equation y = Vmax * x / (Km + x)
Plot reaction rate
Vmax 2.14837E-4 ± 6.90842
Km 1.57029 ± 0.1287
Reduced Chi-S 2.50276E-11
R-Square (COD 0.99174
Adj. R-Square 0.99105

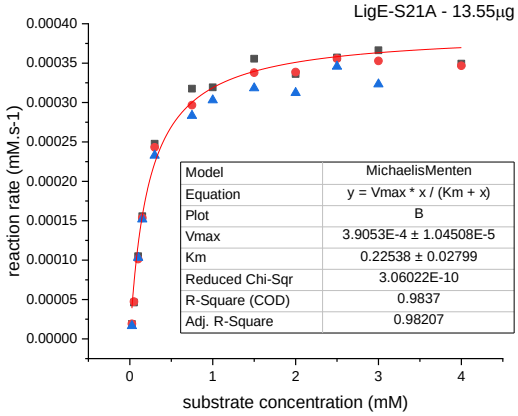

LigE-Y122A - 4.7µg
reaction rate (mM.s-1)
substrate concentration (mM)
Model MichaelisMenten
Equation y = Vmax * x / (Km + x)
Plot reaction rate
Vmax 8.06775E-4 ± 1.63013E-
Km 0.98757 ± 0.04919
Reduced Chi-Sqr 8.509E-11
R-Square (COD) 0.99811
Adj. R-Square 0.99792

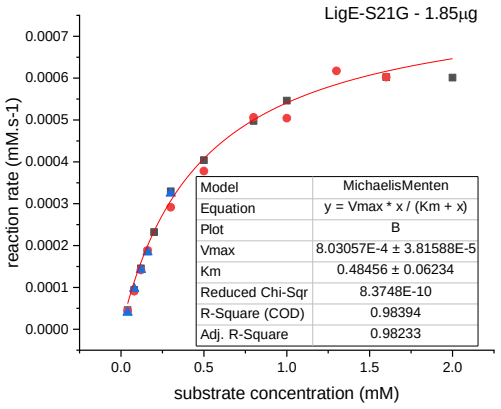

LigE-F142A - 8.6µg
reaction rate (mM.s-1)
substrate concentration (mM)
Model MichaelisMenten
Equation y = Vmax * x / (Km + x)
Plot reaction rate
Vmax 5.17002E-4 ± 2.59384E-5
Km 3.62749 ± 0.34108
Reduced Chi-Sqr 7.23937E-11
R-Square (COD) 0.99415
Adj. R-Square 0.9937

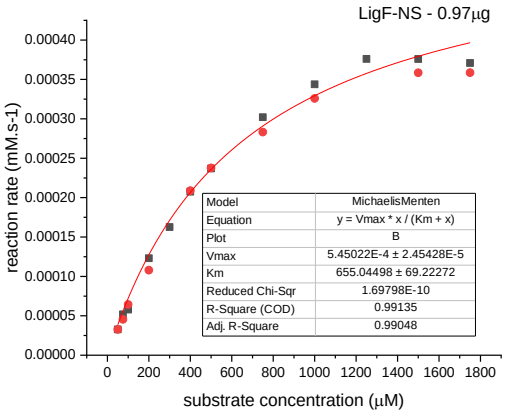

LigE-W197A - 3.2µg
reaction rate (mM.s-1)
substrate concentration (mM)
Model SubstrateInhib
Equation y = Vmax * x / (Km + x * (1 + x / Ki));
Plot reaction rate
Vmax 4.89782E-4 ± 7.03517E-5
Km 1.70273 ± 0.38442
Ki 7.30221 ± 2.40553
Reduced Chi-Sqr 6.05103E-11
R-Square (COD) 0.99251
Adj. R-Square 0.99084

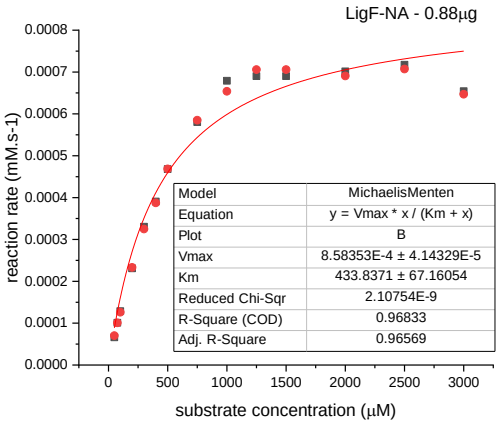

LigE-Y23F - 24.51µg
reaction rate (mM.s-1)
substrate concentration (mM)
Model MichaelisMenten
Equation y = Vmax * x / (Km +
Plot reaction rate
Vmax 0.00133 ± 4.37191E-
Km 1.48903 ± 0.11608
Reduced Chi-Sq 3.96286E-10
R-Square (COD) 0.99736
Adj. R-Square 0.99706

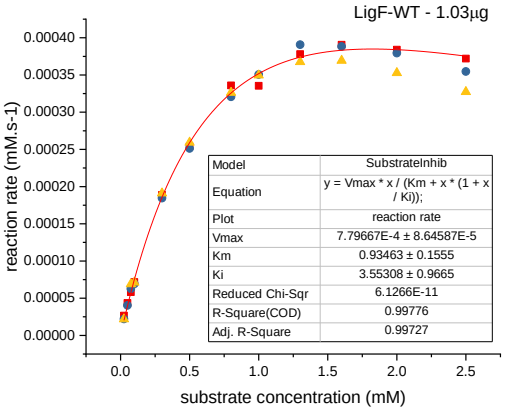

LigE-Y122F - 0.36µg
reaction rate (mM.s-1)
substrate concentration (mM)
Model MichaelisMenten
Equation y = Vmax * x / (Km + x)
Plot reaction rate
Vmax 2.14007E-4 ± 2.421E-6
Km 0.11194 ± 0.00536
Reduced Chi-Sqr 1.26599E-11
R-Square (COD) 0.9963
Adj. R-Square 0.99577

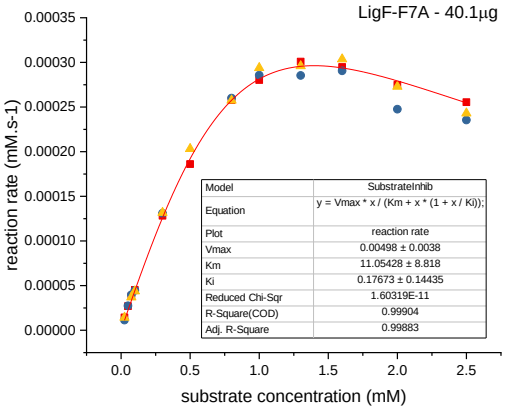

LigE-S21A - 13.55µg
reaction rate (mM.s-1)
substrate concentration (mM)
Model MichaelisMenten
Equation y = Vmax * x / (Km + x)
Plot B
Vmax 3.9053E-4 ± 1.04508E-5
Km 0.22538 ± 0.02799
Reduced Chi-Sqr 3.06022E-10
R-Square (COD) 0.9837
Adj. R-Square 0.98207

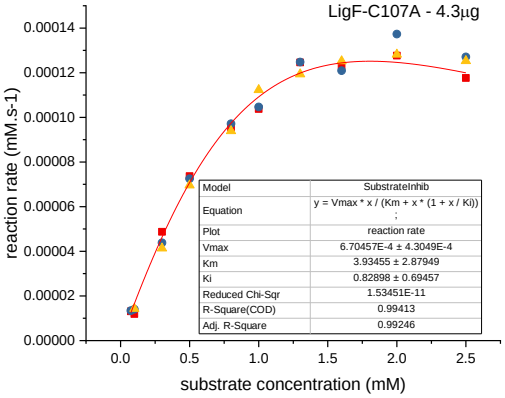

LigE-S21G - 1.85µg
reaction rate (mM.s-1)
substrate concentration (mM)
Model MichaelisMenten
Equation y = Vmax * x / (Km + x)
Plot B
Vmax 8.03057E-4 ± 3.81588E-5
Km 0.48456 ± 0.06234
Reduced Chi-Sqr 8.3748E-10
R-Square (COD) 0.98394
Adj. R-Square 0.98233

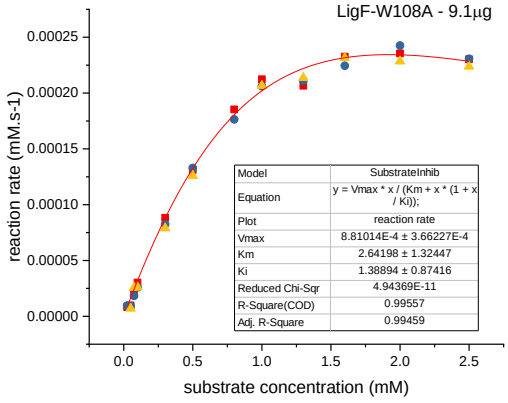

LigF-NS - 0.97µg
reaction rate (mM.s-1)
substrate concentration (µM)
Model MichaelisMenten
Equation y = Vmax * x / (Km + x)
Plot B
Vmax 5.45022E-4 ± 2.45428E-5
Km 655.04498 ± 69.22272
Reduced Chi-Sqr 1.69798E-10
R-Square (COD) 0.99135
Adj. R-Square 0.99048

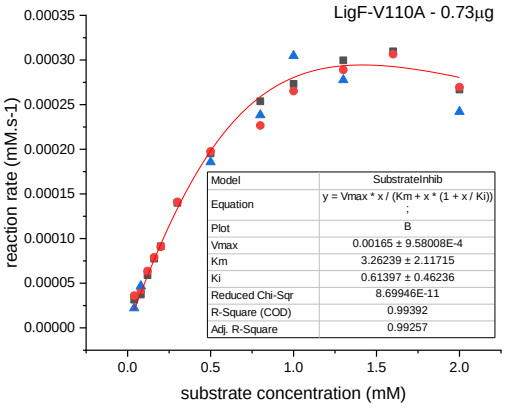

LigF-NA - 0.88µg
reaction rate (mM.s-1)
substrate concentration (µM)
Model MichaelisMenten
Equation y = Vmax * x / (Km + x)
Plot B
Vmax 8.58353E-4 ± 4.14329E-5
Km 433.8371 ± 67.16054
Reduced Chi-Sqr 2.10754E-9
R-Square (COD) 0.96833
Adj. R-Square 0.96569

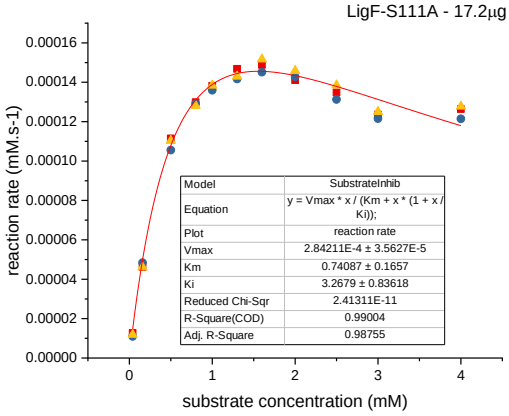

LigF-WT - 1.03µg
reaction rate (mM.s-1)
substrate concentration (mM)
Model SubstrateInhib
Equation y = Vmax * x / (Km + x * (1 + x / Ki));
Plot reaction rate
Vmax 7.79667E-4 ± 8.64587E-5
Km 0.93463 ± 0.1555
Ki 3.55308 ± 0.9665
Reduced Chi-Sqr 6.1266E-11
R-Square(COD) 0.99776
Adj. R-Square 0.99727

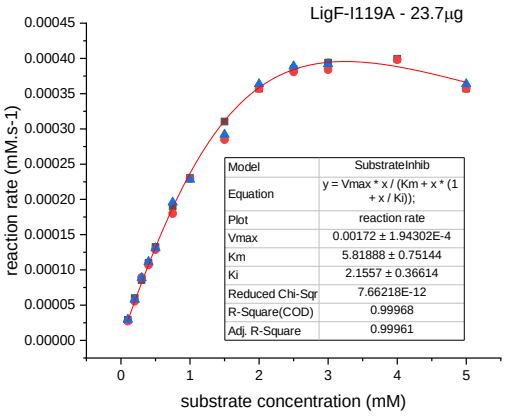

LigF-F7A - 40.1µg
reaction rate (mM.s-1)
substrate concentration (mM)
Model SubstrateInhib
Equation y = Vmax * x / (Km + x * (1 + x / Ki));
Plot reaction rate
Vmax 0.00498 ± 0.0038
Km 11.05428 ± 8.818
Ki 0.17673 ± 0.14435
Reduced Chi-Sqr 1.60319E-11
R-Square(COD) 0.99904
Adj. R-Square 0.99883

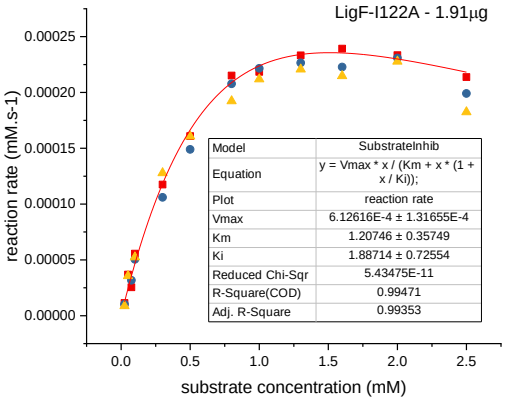

LigF-C107A - 4.3µg
reaction rate (mM.s-1)
substrate concentration (mM)
Model SubstrateInhib
Equation y = Vmax * x / (Km + x * (1 + x / Ki));
Plot reaction rate
Vmax 6.70457E-4 ± 4.3049E-4
Km 3.93455 ± 2.87949
Ki 0.82898 ± 0.69457
Reduced Chi-Sqr 1.53451E-11
R-Square(COD) 0.99413
Adj. R-Square 0.99246

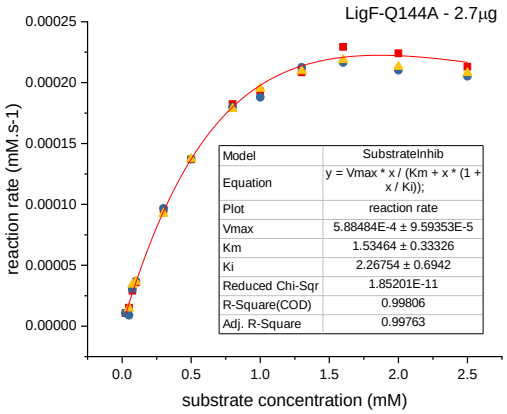

LigF-W108A - 9.1µg
reaction rate (mM.s-1)
substrate concentration (mM)
Model SubstrateInhib
Equation y = Vmax * x / (Km + x * (1 + x / Ki));
Plot reaction rate
Vmax 8.81014E-4 ± 3.66227E-4
Km 2.64198 ± 1.32447
Ki 1.38894 ± 0.87416
Reduced Chi-Sqr 4.94369E-11
R-Square(COD) 0.99557
Adj. R-Square 0.99459

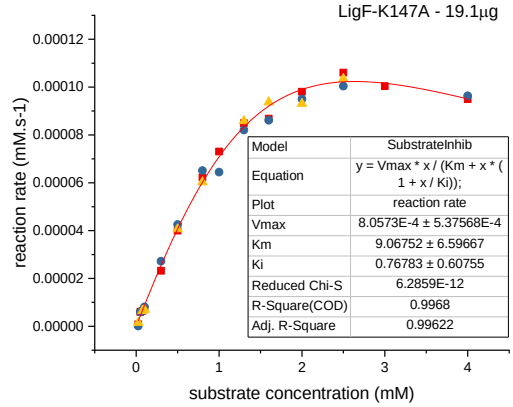

LigF-V110A - 0.73µg
reaction rate (mM.s-1)
substrate concentration (mM)
Model SubstrateInhib
Equation y = Vmax * x / (Km + x * (1 + x / Ki));
Plot B
Vmax 0.00165 ± 9.58008E-4
Km 3.26239 ± 2.11715
Ki 0.61397 ± 0.46236
Reduced Chi-Sqr 8.69946E-11
R-Square (COD) 0.99392
Adj. R-Square 0.99257

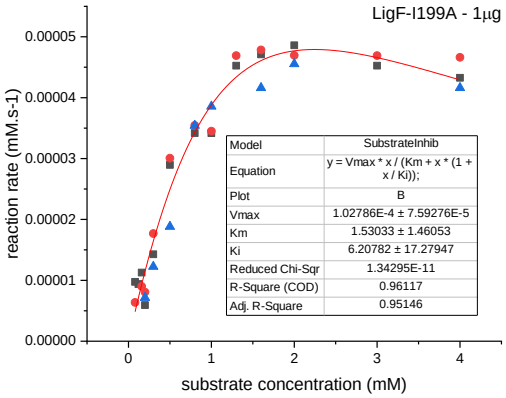

LigF-S111A - 17.2µg
reaction rate (mM.s-1)
substrate concentration (mM)
Model SubstrateInhib
Equation y = Vmax * x / (Km + x * (1 + x / Ki));
Plot reaction rate
Vmax 2.84211E-4 ± 3.5627E-5
Km 0.74087 ± 0.1657
Ki 3.2679 ± 0.83618
Reduced Chi-Sqr 2.41311E-11
R-Square(COD) 0.99004
Adj. R-Square 0.98755

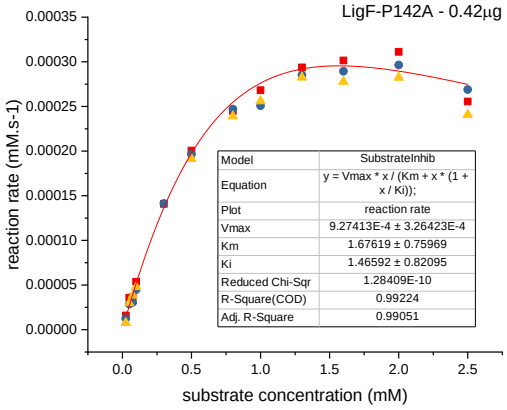

LigF-I119A - 23.7µg
reaction rate (mM.s-1)
substrate concentration (mM)
Model SubstrateInhib
Equation y = Vmax * x / (Km + x * (1 + x / Ki));
Plot reaction rate
Vmax 0.00172 ± 1.94302E-4
Km 5.81888 ± 0.75144
Ki 2.1557 ± 0.36614
Reduced Chi-Sqr 7.66218E-12
R-Square(COD) 0.99968
Adj. R-Square 0.99961

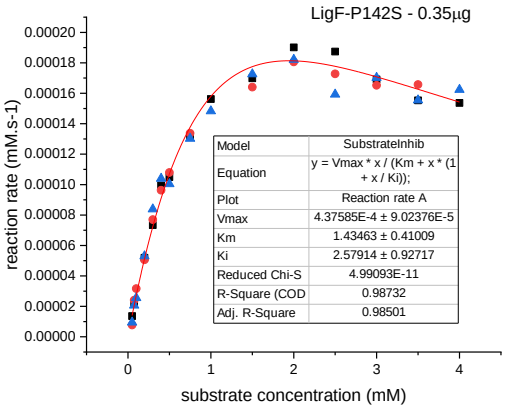

LigF-I122A - 1.91µg
reaction rate (mM.s-1)
substrate concentration (mM)
Model SubstrateInhib
Equation y = Vmax * x / (Km + x * (1 + x / Ki));
Plot reaction rate
Vmax 6.12616E-4 ± 1.31655E-4
Km 1.20746 ± 0.35749
Ki 1.88714 ± 0.72554
Reduced Chi-Sqr 5.43475E-11
R-Square(COD) 0.99471
Adj. R-Square 0.99353

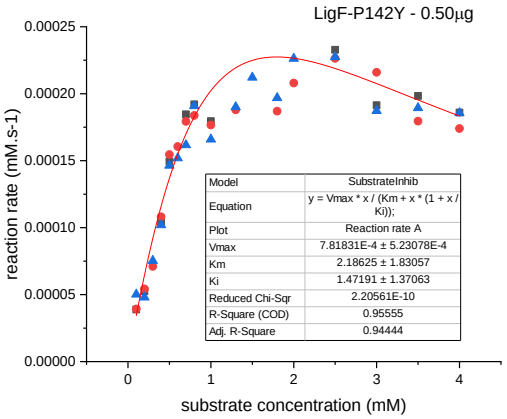

LigF-Q144A - 2.7µg
reaction rate (mM.s-1)
substrate concentration (mM)
Model SubstrateInhib
Equation y = Vmax * x / (Km + x * (1 + x / Ki));
Plot reaction rate
Vmax 5.88484E-4 ± 9.59353E-5
Km 1.53464 ± 0.33326
Ki 2.26754 ± 0.6942
Reduced Chi-Sqr 1.85201E-11
R-Square(COD) 0.99806
Adj. R-Square 0.99763

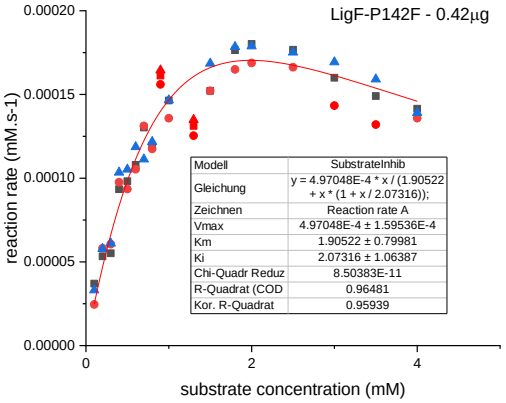

LigF-K147A - 19.1µg
reaction rate (mM.s-1)
substrate concentration (mM)
Model SubstrateInhib
Equation y = Vmax * x / (Km + x * (1 + x / Ki));
Plot reaction rate
Vmax 8.0573E-4 ± 5.37568E-4
Km 9.06752 ± 6.59667
Ki 0.76783 ± 0.60755
Reduced Chi-S 6.2859E-12
R-Square(COD) 0.9968
Adj. R-Square 0.99622

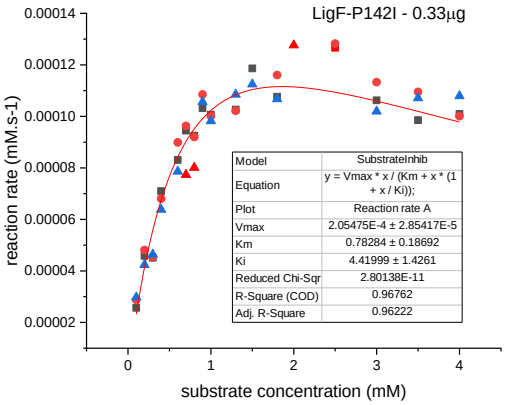

LigF-I199A - 1µg
reaction rate (mM.s-1)
substrate concentration (mM)
Model SubstrateInhib
Equation y = Vmax * x / (Km + x * (1 + x / Ki));
Plot B
Vmax 1.02786E-4 ± 7.59276E-5
Km 1.53033 ± 1.46053
Ki 6.20782 ± 17.27947
Reduced Chi-Sqr 1.34295E-11
R-Square (COD) 0.96117
Adj. R-Square 0.95146

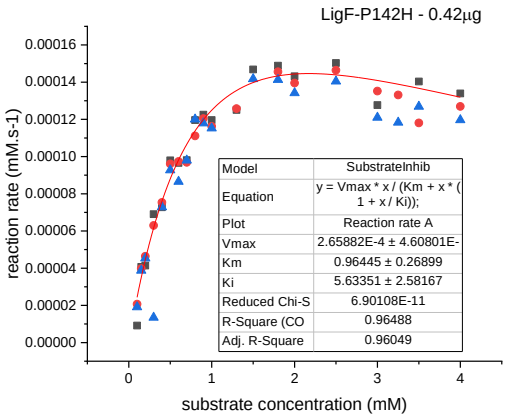

LigF-P142A - 0.42µg
reaction rate (mM.s-1)
substrate concentration (mM)
Model SubstrateInhib
Equation y = Vmax * x / (Km + x * (1 + x / Ki));
Plot reaction rate
Vmax 9.27413E-4 ± 3.26423E-4
Km 1.67619 ± 0.75969
Ki 1.46592 ± 0.82095
Reduced Chi-Sqr 1.28409E-10
R-Square(COD) 0.99224
Adj. R-Square 0.99051

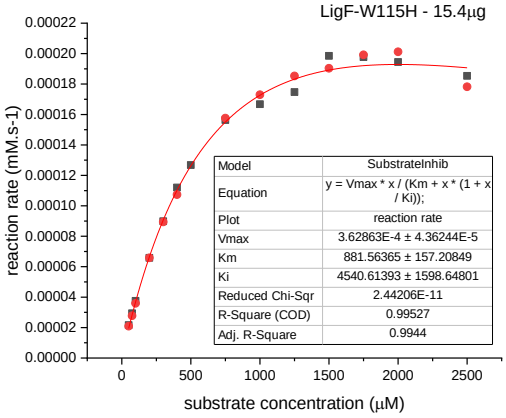

LigF-P142S - 0.35µg
reaction rate (mM.s-1)
substrate concentration (mM)
Model SubstrateInhib
Equation y = Vmax * x / (Km + x * (1 + x / Ki));
Plot Reaction rate A
Vmax 4.37585E-4 ± 9.02376E-5
Km 1.43463 ± 0.41009
Ki 2.57914 ± 0.92717
Reduced Chi-S 4.99093E-11
R-Square (COD 0.98732
Adj. R-Square 0.98501

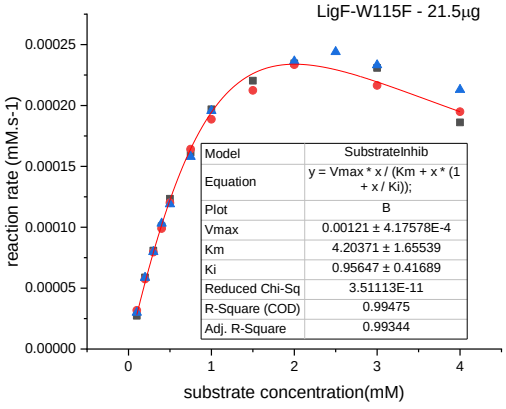

LigF-P142Y - 0.50µg
reaction rate (mM.s-1)
substrate concentration (mM)
Model SubstrateInhib
Equation y = Vmax * x / (Km + x * (1 + x / Ki));
Plot Reaction rate A
Vmax 7.81831E-4 ± 5.23078E-4
Km 2.18625 ± 1.83057
Ki 1.47191 ± 1.37063
Reduced Chi-Sqr 2.20561E-10
R-Square (COD) 0.95555
Adj. R-Square 0.94444

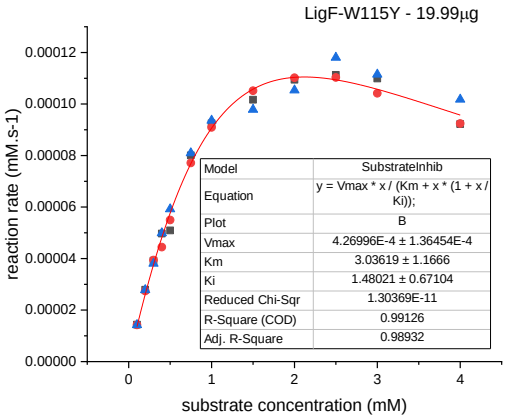

LigF-P142F - 0.42µg
reaction rate (mM.s-1)
substrate concentration (mM)
Modell SubstrateInhib
Gleichung y = 4.97048E-4 * x / (1.90522 + x * (1 + x / 2.07316));
Zeichnen Reaction rate A
Vmax 4.97048E-4 ± 1.59536E-4
Km 1.90522 ± 0.79981
Ki 2.07316 ± 1.06387
Chi-Quadr Reduz 8.50383E-11
R-Quadrat (COD 0.96481
Kor. R-Quadrat 0.95939

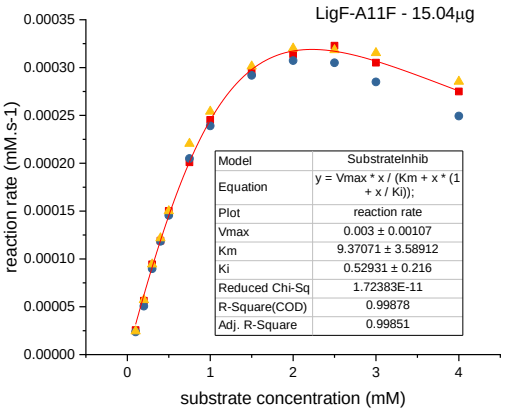

LigF-P142I - 0.33µg
reaction rate (mM.s-1)
substrate concentration (mM)
Model SubstrateInhib
Equation y = Vmax * x / (Km + x * (1 + x / Ki));
Plot Reaction rate A
Vmax 2.05475E-4 ± 2.85417E-5
Km 0.78284 ± 0.18692
Ki 4.41999 ± 1.4261
Reduced Chi-Sqr 2.80138E-11
R-Square (COD) 0.96762
Adj. R-Square 0.96222

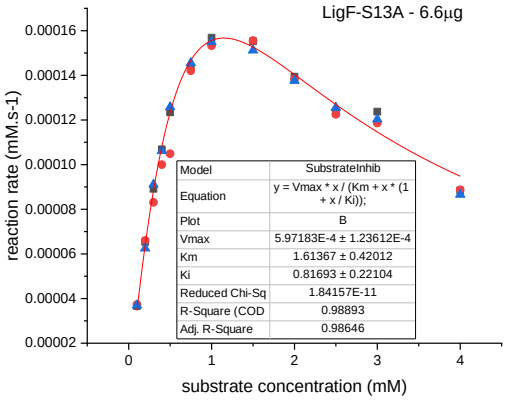

LigF-P142H - 0.42µg
reaction rate (mM.s-1)
substrate concentration (mM)
Model SubstrateInhib
Equation y = Vmax * x / (Km + x * (1 + x / Ki));
Plot Reaction rate A
Vmax 2.65882E-4 ± 4.60801E-
Km 0.96445 ± 0.26899
Ki 5.63351 ± 2.58167
Reduced Chi-S 6.90108E-11
R-Square (CO 0.96488
Adj. R-Square 0.96049

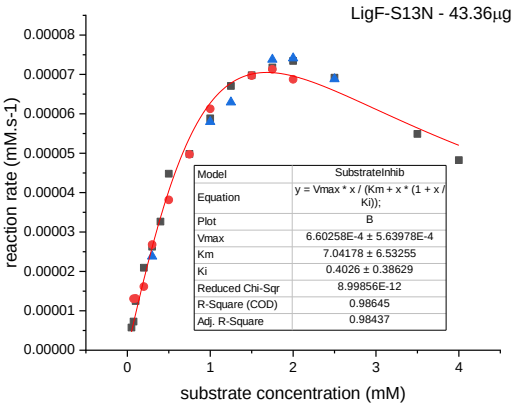

LigF-W115H - 15.4µg
reaction rate (mM.s-1)
substrate concentration (µM)
Model SubstrateInhib
Equation y = Vmax * x / (Km + x * (1 + x / Ki));
Plot reaction rate
Vmax 3.62863E-4 ± 4.36244E-5
Km 881.56365 ± 157.20849
Ki 4540.61393 ± 1598.64801
Reduced Chi-Sqr 2.44206E-11
R-Square (COD) 0.99527
Adj. R-Square 0.9944

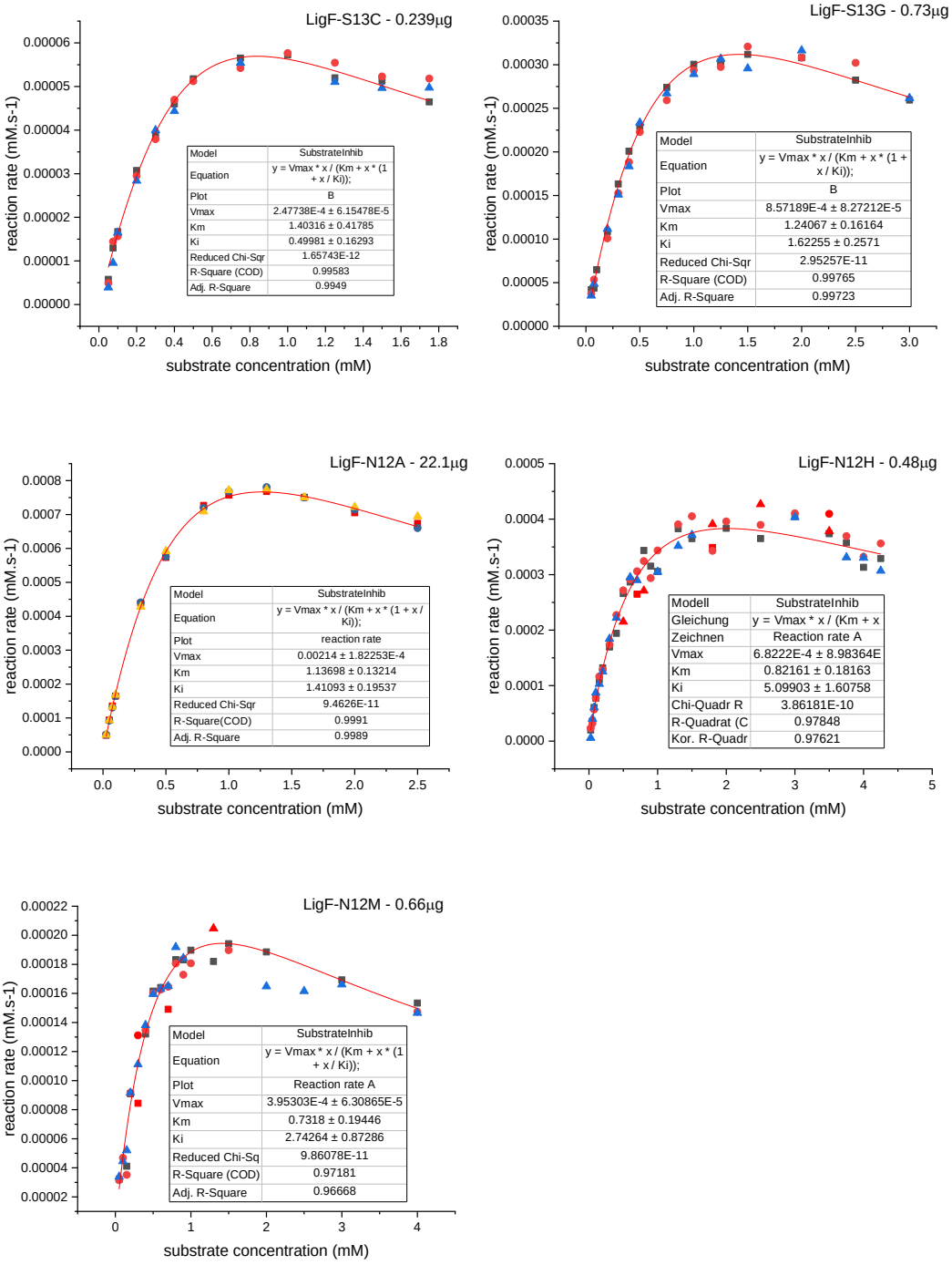

LigF-W115F - 21.5µg
reaction rate (mM.s-1)
substrate concentration(mM)
Model SubstrateInhib
Equation y = Vmax * x / (Km + x * (1 + x / Ki));
Plot B
Vmax 0.00121 ± 4.17578E-4
Km 4.20371 ± 1.65539
Ki 0.95647 ± 0.41689
Reduced Chi-Sq 3.51113E-11
R-Square (COD) 0.99475
Adj. R-Square 0.99344


LigF-W115Y - 19.99µg
reaction rate (mM.s-1)
substrate concentration (mM)
Model SubstrateInhib
Equation y = Vmax * x / (Km + x * (1 + x / Ki));
Plot B
Vmax 4.26996E-4 ± 1.36454E-4
Km 3.03619 ± 1.1666
Ki 1.48021 ± 0.67104
Reduced Chi-Sqr 1.30369E-11
R-Square (COD) 0.99126
Adj. R-Square 0.98932


LigF-A11F - 15.04µg
reaction rate (mM.s-1)
substrate concentration (mM)
Model SubstrateInhib
Equation y = Vmax * x / (Km + x * (1 + x / Ki));
Plot reaction rate
Vmax 0.003 ± 0.00107
Km 9.37071 ± 3.58912
Ki 0.52931 ± 0.216
Reduced Chi-Sq 1.72383E-11
R-Square(COD) 0.99878
Adj. R-Square 0.99851


LigF-S13A - 6.6µg
reaction rate (mM.s-1)
substrate concentration (mM)
Model SubstrateInhib
Equation y = Vmax * x / (Km + x * (1 + x / Ki));
Plot B
Vmax 5.97183E-4 ± 1.23612E-4
Km 1.61367 ± 0.42012
Ki 0.81693 ± 0.22104
Reduced Chi-Sq 1.84157E-11
R-Square (COD 0.98893
Adj. R-Square 0.98646


LigF-S13N - 43.36µg
reaction rate (mM.s-1)
substrate concentration (mM)
Model SubstrateInhib
Equation y = Vmax * x / (Km + x * (1 + x / Ki));
Plot B
Vmax 6.60258E-4 ± 5.63978E-4
Km 7.04178 ± 6.53255
Ki 0.4026 ± 0.38629
Reduced Chi-Sqr 8.99856E-12
R-Square (COD) 0.98645
Adj. R-Square 0.98437



Supplementary Figure 7.15: Kinetic analyses of every enzyme and mutant created in this project toward substrate **3**.

7.9 LigE homologs multiple sequence alignment

```
                                          10         20         30         40
                                   ....|....| ....|....| ....|....| ....|....|
Sphingobium sp. SYK-6(1)           MARNNTITLY DLQLESGCTI SPYVWRTKYA LKHKGFDIDI
Novosphingobium fuchskuhlense      MAANNTITFY DLALSTGATI SPFVWATKYA LKHKGFDLDV
Novosphingobium sp. NDB2Meth1      MAANNTITFY DLALSTGATI SPFVWATKYA LKHKGFDLDV
Novosphingobium sp. AAP93          MAANNTITFY DLALSTGATI SPFVWATKYA LKHKGFDLDV
Novosphingobium subterraneum       MAANNTITFY DLALSTGATI SPFVWATKYA LKHKGFDLDV
Novosphingobium sp. CCH12-A3       MAANNTITFY DLALSTGATI SPFVWATKYA LKHKGFDLDV
Novosphingobium sp. B1             MAANNTITFY DLALSTGATI SPFVWATKYA LKHKGFDLDV
Novosphingobium sp. AAP83          MAQNNTITFY DLALSTGATI SPFVWATKYA LKHKGFDLDV
Novosphingobium aromaticivorans    MAANNTITFY DLALSTGATI SPFVWATKYA LKHKGFDLDV
Novosphingobium sp. B-7            MAQDNKITFY DLALSTGATI SPFVWATKYA LKHKGFDLDV
Novosphingobium sp. AAP1           MAQDNKITFY DLALSTGATI SPFVWATKYA LKHKGFDLDV
Altererythrobacter atlanticus(1)   MAKNNTITFY DLALSTGATI SPFVWATKYA LKHKGFDLDV
Erythrobacter sp. SG61-1L(1)       MAKNNTITFY DLALSTGATI SPFVWATKYA LKHKGFDLDV
Novosphingobium capsulatum         MAQDNKITFY DLAISTGATI SPFVWATKYA LKHKGFDLDV
Novosphingobium sp. SCN 66-18      MAANNTITFY DLALSTGATI SPFVWATKYA VKHKGFELDV
Novosphingobium sp. MD-1           MAANNTITFY DLALSTGATI SPFVWATKYA VKHKGFELDV
Altererythrobacter sp. 66-12       MARNNTITFY DLALSTGATI SPFVWATKYA LKHKGFDLDV
Sphingomonadales bacterium 63-6(1) MAKNNTITFY DLALSTGATI SPFVWATKYA LKHKGFDLDV
Altererythrobacter sp. Root672(1)  MAQNNTITFY DLAISTGATI SPFVWATKYA LKHKGFDLDV
Sphingomonadales bacterium 63-6(2) MAKNNRITLY DLQIAAGCTI SPFVWATKYA VAHKGFEIDI
Erythrobacter sp. SG61-1L(2)       MAKNNTITLY DLQLEPGCTI SPFVWATKYA IAHKGFEIDI
Novosphingobium acidiphilum        MAKDNRITFF DLQHASGATT SPFVWATKYA LKHKGFDLDV
Novosphingobium sp. Fuku2-ISO-50   MAKDNRIRFF DLQHASGCTT SPFVWATKYA LKHKGFDLDV
Sphingobium sp. 66-54(1)           MAKDNKITIY DLALASGATI SPFVWATKYA IAHKGFELDI
Altererythrobacter sp. Root672(2)  MAKDNTITLY DLQLASGATI SPFVWATKYA IAHKGFEMDI
Altererythrobacter atlanticus(2)   MAANNTVTLY DLQLASGATI SPFVWATKLA IAHKGLDMEI
Novosphingobium sp. SCN 63-17      MAKDNKITFF DLTHESGCTT SPFVWATKYA VKHKGFDLDV
Sphingobium sp. SCN 64-10          MAKDNKITIY DLALASGATI SPFVWATKYA IAHKGFELDI
Sphingobium sp. SYK-6(2)           MAKDNKITIY DLALASGATI SPFVWATKYA IAHKGFELDI
Novosphingobium sp. FSW06-99       MAQNNRITFF DLQHASGATT SPFVWATKYA LKHKGFDLDV
Sphingomonas hengshuiensis         MTDRHSVTLY DLNLASGCTI SPFVWRTKYA LAHKGLAIDI
Novosphingobium sp. PP1Y           MAKDNRITLY DLQLASGCTI SPFVWRTKYA LAHKGFDVDI
Novosphingobium sp. ST904          MARDNRITLY DLQLASGCTI SPFVWRTRYA LAHKGFDVDL
Novosphingobium mathurense         MARDNRITLY DLQLASGCTI SPFVWRTKYA LAHKGFDIDI
Novosphingobium sp. KN65.2         MARDNRITLY DLQLASGCTI SPFVWRSKYA LAHKGFDIDI
Sphingobium sp. 66-54(2)           MAKNNKITLF DLQLESGCTI SPYVWRTKYA LAHKGFDVEL
Novosphingobium sp. MBES04         MAKDNRITLY DLQLASGCTI SPFVWRTKYA LAHKGFDMDI

                                          50         60         70         80
                                   ....|....| ....|....| ....|....| ....|....|
Sphingobium sp. SYK-6(1)           VPGGFTGILE RTGGRSERVP VIVDDGEWVL DSWVIAEYLD
Novosphingobium fuchskuhlense      VPGGFTGILE RTGGKTERLP AIVDDGEWVL DSWGIVEYLD
Novosphingobium sp. NDB2Meth1      VPGGFTGILE RTGGKTERLP AIVDDGEWVL DSWGIVEYLD
Novosphingobium sp. AAP93          VPGGFTGILE RTGGKTERLP AIVDDGEWVL DSWGIVEYLD
Novosphingobium subterraneum       VPGGFTGILE RTGGKTERLP AIVDDGEWVL DSWGIVEYLD
Novosphingobium sp. CCH12-A3       VPGGFTGILE RTGGKTERLP AIVDDGEWVL DSWGIVEYLD
Novosphingobium sp. B1             VPGGFTGILE RTGGKTERLP AIVDDGEVVL DSWGIVEYLD
Novosphingobium sp. AAP83          VPGGFTGILE RTGGKTERLP AIVDDGEFVL DSWGIVEYLD
Novosphingobium aromaticivorans    VPGGFTGILE RTGGKTERLP AIVDDGEFVL DSWGIVEYLD
Novosphingobium sp. B-7            VPGGFTGILE RTGGKTERLP AIVDDGTWVL DSWGIVEYLD
Novosphingobium sp. AAP1           VPGGFTGILE RTGGKTERLP AIVDDGTWVL DSWGIVEYLD
Altererythrobacter atlanticus(1)   VPGGFTGIPE RTGGKTERLP AIVDDGKWVL DSWGIVEYLD
Erythrobacter sp. SG61-1L(1)       VPGGFTKILE RTGGKTERLP AIVDDGTWVL DSWGIVEYLD
Novosphingobium capsulatum         VPGGFTGILE RTGGKTERLP AIVDDGTWVL DSWGIVEYLD
Novosphingobium sp. SCN 66-18      VPGGFTGILE RTGGKTERLP AIVDDGEWVL DSWGIVEYLD
Novosphingobium sp. MD-1           VPGGFTGILE RTGGKTERLP AIVDDGEWVL DSWGIVEYLD
Altererythrobacter sp. 66-12       VPGGFTKIPE RTGGKTERLP AIVDDGKWVL DSWGIVEYLD
Sphingomonadales bacterium 63-6(1) VPGGFTGIPE RTGGNSERLP AIVDDGKWVL DSWGIVEYLD
Altererythrobacter sp. Root672(1)  VPGGFTKIPE RTGGVTERLP AIVDDGKWVL DSWGIVEYLD
Sphingomonadales bacterium 63-6(2) VPGGFTGIME RTGGKTERLP AIVDDGEWVL DSWLIAEYLD
```

```
Erythrobacter sp. SG61-1L(2)          VPGGFTGIME RTGGKTERLP AIVDDGEWVL DSWLIAEYLD
Novosphingobium acidiphilum           VDGGFTGILE RTQGRSERLP VIVDDGEWVL DSWLIAEYLD
Novosphingobium sp. Fuku2-ISO-50      VDGGFTGILD RTGGRSERLP VIVDDGEWVL DSWLIAEYLD
Sphingobium sp. 66-54(1)              VPGGFSGIPE RTGGKTERLP AIVDDGKWVL DSWLIAEYLD
Altererythrobacter sp. Root672(2)     VPGGFTGIPE RTGGQTERLP AIVDDGQWVL DSWLIAEYLD
Altererythrobacter atlanticus(2)      VPGGFTGIEE RTGGKTQRLP AIVDDGEWIL DSWTIAEYLD
Novosphingobium sp. SCN 63-17         VPGGFTGILD RTGGRSERLP VICDDGEYVL DSWLIAEYLD
Sphingobium sp. SCN 64-10             VPGGFSGIPE RTGGVTERLP AIVDDGKWVL DSWLIAEYLD
Sphingobium sp. SYK-6(2)              VPGGFSGIPE RTGGVTERLP AIVDDGKWVL DSWLIAEYLD
Novosphingobium sp. FSW06-99          IDGGFIGILD RTGGRSERLP VIVDDGEWIL DSWLIAEHLD
Sphingomonas hengshuiensis            VPGGFTGILE RTGGRSERLP AIIDNGEWVL DSWLIAEYLD
Novosphingobium sp. PP1Y              VPGGFTGIAE RTGGRSERVP VIVDDGEWVL DSWKIAEYLD
Novosphingobium sp. ST904             VPGGFTGIAE RTGGRSERVP VIVDDGEWVL DSWKIAEYLD
Novosphingobium mathurense            VPGGFTGIAE RTGGRSERVP VIVDDGEWVL DSWKIAEYLD
Novosphingobium sp. KN65.2            VPGGFTGIAE RTGGRSERVP VIVDDGEWVL DSWKIAEYLD
Sphingobium sp. 66-54(2)              VPGGFTGILE RTGGRSERVP VIVDDGEWIL DSWVIAEYLD
Novosphingobium sp. MBES04            VPGGFTGIAE RTGGRSERAP VIVDDGKWVL DSWKIAEYLD

                                                90        100       110       120
                                      ....|....| ....|....| ....|....| ....|....|
Sphingobium sp. SYK-6(1)              EKYPDRPMLF EGPTQKNLMK FLDNWLWSTA VGPWFRCYIL
Novosphingobium fuchskuhlense         AKYPDRPALI PHPSVAATLK ALDHWFWNAA VGPWMFCFCA
Novosphingobium sp. NDB2Meth1         AKYPDRPALI PHPSVAATLK ALDHWFWGAA VGPWMRCFCA
Novosphingobium sp. AAP93             AKYPDRTALI PHPSVAATLK ALDHWFWGAA VGPWMRCFCA
Novosphingobium subterraneum          AKYPNRPALI PHESVAVTLK ALDHWFWNAA VGPWMFCFCA
Novosphingobium sp. CCH12-A3          AKYPDRPALI PHESVAVTLK ALDHWFWNAA VGPWMFCFCA
Novosphingobium sp. B1                AKYPDRPALI PHESVAATLK ALDHWFWNAA VGPWMFCFCA
Novosphingobium sp. AAP83             AKYPDRPALI PHESVAATLK ALDHWFWNTA VGPWMFCFCQ
Novosphingobium aromaticivorans       AKYPDRPVLI PHESVAATLK ALDNWFWNAA VGPWMFCFCQ
Novosphingobium sp. B-7               KTYPDRPLLI PHESVAATLK ALDHWFWGAA VGPWMRCFCA
Novosphingobium sp. AAP1              KTYPDRPLLI PHESVAATLK ALDHWFWGAA VGPWMRCFCA
Altererythrobacter atlanticus(1)      ETYPDRPALI PHPSVAALTR AMDAWFWKVA TGPWMRCFCA
Erythrobacter sp. SG61-1L(1)          AEYPTRPTLI PHESVATVTR ALDAWFWKVA TGPWMRCFCA
Novosphingobium capsulatum            ATYPDRPLLI PHESVAATLK ALDHWFWGAA VGPWMRCFCA
Novosphingobium sp. SCN 66-18         ATYPDRPALI PHESVAATLK ALDHWFWNAA VGPWMFCFCA
Novosphingobium sp. MD-1              AKYPDRPALI PHESVAATLK ALDHWFWNAA VGPWMFCFCA
Altererythrobacter sp. 66-12          ETYPDRPMLI PHPSVAAVTR ALDAWFWGAA TGPWMRCFCV
Sphingomonadales bacterium 63-6(1)    AEYPARPTLV PHESVASVTR ALDAWFWKVA TGPWMRCFCA
Altererythrobacter sp. Root672(1)     ATYPDRPLLI PHPSVAIVTR ALDAWFWQVA TGPWMRCNCV
Sphingomonadales bacterium 63-6(2)    EKYPDRPTLI PDPSLRVLTT MMESWLWQVA IGPWMSCFIK
Erythrobacter sp. SG61-1L(2)          EKYPDRPTLI PDPSVRVLTT MMESWLWQAA IGPWMTCFIK
Novosphingobium acidiphilum           RKYPDRPTLI GDPSVRVLAQ FLETWLWKTA VGPWARCFAV
Novosphingobium sp. Fuku2-ISO-50      AKYPDRPTLI GDPSVKVLTQ FLEAWLWKTA VGPWARCFAV
Sphingobium sp. 66-54(1)              ETYPDRPTLI PHPSVKALTQ AMEGWLWQTA ISPWMTCFIK
Altererythrobacter sp. Root672(2)     EKYPHRPTLI GDPSVKVCGQ MLEQWLWQTA IGPWMTCYLK
Altererythrobacter atlanticus(2)      EKYPDRPTLI GDPGIRPSAQ MNEAWLWQTA VGPWMTCYLV
Novosphingobium sp. SCN 63-17         AKYPDRPTLI GDPSVKVLTQ FLETWLWKTV VGPWARCFAV
Sphingobium sp. SCN 64-10             ETYPDRPTLI PHPSVKALTQ GMEAWLWSTA ISPWMTCFIK
Sphingobium sp. SYK-6(2)              ETYPERPTLI PHASVKALTQ GMEAWLWGAA ISPWMTCFIK
Novosphingobium sp. FSW06-99          RKYPDRPTLI GDPSVRVLTQ FLETWLWKTA VGPWARCFAV
Sphingomonas hengshuiensis            GKYPERP~LF EGPSMKVLTK FIDQWLWRTA IGPWFRCYIL
Novosphingobium sp. PP1Y              ERYPDRPMLF EGPSMKQLTK FLDAWLWQTA IGPWFRCYIQ
Novosphingobium sp. ST904             EKYPDRPMLF EGPSMKELTR FLDAWLWTTA IGPWFSCYIL
Novosphingobium mathurense            ERYPERPMLF EGPSMQQLTK FLDAWLWQTA IGPWFRCYIQ
Novosphingobium sp. KN65.2            ERYPERPMLF EGPSMQQLTK FLDAWLWQTA IGPWFRCYIQ
Sphingobium sp. 66-54(2)              EKYPDRPMLF EGPAQKNLMK FLDNWLWGTA IGPWFRCYIL
Novosphingobium sp. MBES04            ETYPDRPMLF EGPSMKVLTK FLDAWLWKTI IAPWFRCYIL

                                                130       140       150       160
                                      ....|....| ....|....| ....|....| ....|....|
Sphingobium sp. SYK-6(1)              DYHDLSLPQD RDYVRWSREQ WFLGG~QRLE DVQAGREDRL
Novosphingobium fuchskuhlense         DYRDLSVPQD HEYVTHSREK ML~~G~RKLE DVQAGREDRL
Novosphingobium sp. NDB2Meth1         DYRDLSVPQD HEYITHSREK ML~~G~KKLE DMQAGREQRL
Novosphingobium sp. AAP93             DYRDLSVPQD HEYITHSREK ML~~G~KKLE DMQAGREQRL
Novosphingobium subterraneum          DYRDLSVPAD HEYVTHSREK ML~~G~RKLE DVQAGREERL
Novosphingobium sp. CCH12-A3          DYRDLSVPAD HEYVTHSREK ML~~G~RKLE DVQAGREERL
Novosphingobium sp. B1                DYRDLSVPAD HEYVTHSREK ML~~G~RKLE EVQAGREERL
```

```
Novosphingobium sp. AAP83                DYRDLSLPQD HEYVTHSREK ML~~G~RKLE DVQAGREERL
Novosphingobium aromaticivorans          DYRDLSLPQD HEYVTHSREK ML~~G~RKLE EVQAGREERL
Novosphingobium sp. B-7                  DYRDLSVPQD HEYITHSREK ML~~G~CKLE EIQAGREDRL
Novosphingobium sp. AAP1                 DYRDLSVPQD HEYITHSREK ML~~G~RKLE EIQAGREDRL
Altererythrobacter atlanticus(1)         NYRDLANKED HEYITHSREI ML~~G~KKLE DMQAGYEERL
Erythrobacter sp. SG61-1L(1)             NYRNLANQED HEYITHSREI ML~~G~MKLE DMQAGYEDRL
Novosphingobium capsulatum               DYRDLSLPQD HEYITHSREK ML~~G~RKLE EIQAGREDRL
Novosphingobium sp. SCN 66-18            DYRNLSLPQD HAYVTHSREK ML~~G~RKLE EVQAGREERL
Novosphingobium sp. MD-1                 DYRNLSVPQD HAYVTHSREK ML~~G~RKLE EVQAGREERL
Altererythrobacter sp. 66-12             NYRDLSNPED HEYITYSREK ML~~G~KTLE EMQAGWEDRL
Sphingomonadales bacterium 63-6(1)       NYRNLANPED HEYITHSREI ML~~G~MKLE DMQAGYEDRL
Altererythrobacter sp. Root672(1)        SYRDLSNPED HEYITHSREK ML~~G~KTLE AMQEGYEDRL
Sphingomonadales bacterium 63-6(2)       SYRDRSLPQD HEYVTESRER MF~~G~RKME DIIVGREDRL
Erythrobacter sp. SG61-1L(2)             SYRDRSLPQD HEYVTQSRER MF~~G~RKME DIIVGREDRL
Novosphingobium acidiphilum              QYRDRSFPQD IPYIVESRRR MW~~G~APME DLVSGREDRL
Novosphingobium sp. Fuku2-ISO-50         GYRDRCFAQD IPYIVESRRR MW~~G~APME DLAAGREDRL
Sphingobium sp. 66-54(1)                 QYRDRSLPQD HAYVTESRER MF~~G~RKIE DIIVGREDRL
Altererythrobacter sp. Root672(2)        QYRDRALPQD HGYVTESRER MF~~GGQKIE DIIIGREDRL
Altererythrobacter atlanticus(2)         AYRDRSVPED HEYVTATRET MF~~GGRKLE DIIVGREDRL
Novosphingobium sp. SCN 63-17            QYRDRCFPHD IQYITESRLR MW~~G~KPME ELIVGREDVF
Sphingobium sp. SCN 64-10                QYRDRSLPVD HEYVTTSRER MF~~G~RKIE DLIVGREDRL
Sphingobium sp. SYK-6(2)                 QYRDRSLPQD HEYVTTSRER MF~~G~RKIE DIIVGREDRI
Novosphingobium sp. FSW06-99             QYRDRCFEQD IPYIVESRRR MW~~G~APME DLIVGREDRL
Sphingomonas hengshuiensis               DYHNLSYPHD HDYIRTTRET MFLGG~QKLE DVQAGREERL
Novosphingobium sp. PP1Y                 DYHDLSLPQD QTYVRHSRET MFLGG~KTLE EVQAGREDRL
Novosphingobium sp. ST904                DYHDLSLPQD HAYVRHSRET MFLGG~RKLE DVQAGREDRL
Novosphingobium mathurense               DYHDLSLPQD QAYVRHSRET MFLGG~KTLE EVQAGREDRL
Novosphingobium sp. KN65.2               DYHDLSLPQD QAYVRHSRET MFLGG~KTLE EVQAGREDRL
Sphingobium sp. 66-54(2)                 DYHDLSLPQD RDYVRWSREQ WFLGG~QKLE DVQAGREDRL
Novosphingobium sp. MBES04               DYHDLSLPQD HAYVRESRET MFLGG~QKLE DVQAGREDRL

                                                  170       180       190       200
                                         ....|....| ....|....| ....|....| ....|....|
Sphingobium sp. SYK-6(1)                 PLVPPTLEPF RRILAETKWL GGDQPNFADY SALAVFLWTA
Novosphingobium fuchskuhlense            PKISAALEPL RAALGQSEWL GGDSPNYADY RILGGILFTA
Novosphingobium sp. NDB2Meth1            PGISAALEPL RATLGQHEWL GGDAPNYADY RIMGGILFTS
Novosphingobium sp. AAP93                PGISAALEPL RATLGQHEWL GGDAPNYADY RIMGGILFTS
Novosphingobium subterraneum             PRISAALEPL RAALGQHEWL GGSSPNYADF RIMGGILFTA
Novosphingobium sp. CCH12-A3             PRISAALEPL RAALGQHEWL GGSSPNYADF RIMGGILFTA
Novosphingobium sp. B1                   PRISAALEPL RAALGQHEWL GGSSPNYADF RIMGGILFTA
Novosphingobium sp. AAP83                PKISAALEPL RAALAQHQWL GGSTPNYADF RIMGGILFTA
Novosphingobium aromaticivorans          PKISAALEPL RAALAQHQWL GGSSPNYADY RIMGGILFTA
Novosphingobium sp. B-7                  AGISAALEPL RSALGQHAWL GGSSPNYADY RILGGFLFTA
Novosphingobium sp. AAP1                 AGISAALEPL RSALGQHAWL GGSSPNYADY RILGGFLFTA
Altererythrobacter atlanticus(1)         PGISADLEPL RIALREVEWL GGDGPNYADY RIMGSILFTA
Erythrobacter sp. SG61-1L(1)             PQISADLEPL RIALRESKWL GGSQPNYADF RIMGSILFTA
Novosphingobium capsulatum               AGISAALEPL RSALGQHAWL GGSSPNYADY RILGGFLFTA
Novosphingobium sp. SCN 66-18            PKISAALEPL RAALAQHAWL GGATPNYADF RIMGGILFTA
Novosphingobium sp. MD-1                 PKISAALEPL RAALAQHPWL GGPTPNYADF RIMGGILFTA
Altererythrobacter sp. 66-12             PGISAALEPL RIALREVDYL GGDAPNYADY RILGSILFTA
Sphingomonadales bacterium 63-6(1)       PKISAELEPL RIALRESKWL GGSQPNYADY RILGSILFTA
Altererythrobacter sp. Root672(1)        PAISAALEPL RIALREGDWL GGDSPNYADY RILGSILFTA
Sphingomonadales bacterium 63-6(2)       PLVPPSLQLM RNALADNPWF GGESPNYADY RMLSVFLFAA
Erythrobacter sp. SG61-1L(2)             PLVPPALQLL RNTLADNKWF GGESPNYADH RMLSVFLFAA
Novosphingobium acidiphilum              PLVLPELELL RGILREHKWL GGDSPNYADY RALAVFLWCS
Novosphingobium sp. Fuku2-ISO-50         PLVLPELELL RGILRDHKWL GGDAPNYADY RALAVFLWCA
Sphingobium sp. 66-54(1)                 PLVPPTLQLM RNALAENKWF GGDTPNYADF RLLAVFLFTA
Altererythrobacter sp. Root672(2)        PQIPPKLELM RNTLREHQWF GGESPNYVDY RLLAVFLFLA
Altererythrobacter atlanticus(2)         PRISADLELM RGVLRENKWF GGDSPNYADH RMLACFLWLA
Novosphingobium sp. SCN 63-17            PKVLPELELL RGILREHKWL GGETPNYADY RALAVFLWAA
Sphingobium sp. SCN 64-10                PKVPPTLQLR RNVLAENKWF GGETPNYADF RLLAVFLFTA
Sphingobium sp. SYK-6(2)                 PKVPPTLQLL RNVLAENKWL GGDTPNYADF RLLAVFLFTA
Novosphingobium sp. FSW06-99             PLVLPELELL RGILREHKWL GGDTPNYADY RALAVFLWCA
Sphingomonas hengshuiensis               PLVPPLLEPL RQLLRDTPWL GGDTPNYADY CALAVFLWTA
Novosphingobium sp. PP1Y                 PLVPPTLEPF RKLLRDTPWL GGDAPNFADY TALAVFLWTA
Novosphingobium sp. ST904                PLVPPTLEPF RKLLRDTPWL GGEKPNFADY TALAVFLWTA
Novosphingobium mathurense               PLVPPTLEPF RKLLRDTPWL GGDAPNFADY TALSVFLWTA
```

```
Novosphingobium sp. KN65.2          PLVPPTLEPF RKLLRDTPWL GGDAPNFADY TALSVFLWTA
Sphingobium sp. 66-54(2)            PLVPPTLEPF RKILAETKWL GGETPNFADY SALAVFLWTA
Novosphingobium sp. MBES04          PHVPPLLEPL RQLLRDTPWL GGATPNYADY TALAIFLWTG

                                              210        220        230        240
                                    ....|....| ....|....| ....|....| ....|....|
Sphingobium sp. SYK-6(1)            SVART~PPLT EDDPLRDWLD RGFDLFDGLG RHPGMNPLFG
Novosphingobium fuchskuhlense       SVCKT~PVLA EDDPLRGWIE RCLDLYGGLG RHPGLFPLFG
Novosphingobium sp. NDB2Meth1       SVCKV~PVLA NDDPLRGWIE RCLDLYGGLG RHPGLFPLFG
Novosphingobium sp. AAP93           SVCKV~PVLA NDDPLRGWIE RCLDLYGGLG RHPGLFPLFG
Novosphingobium subterraneum        SVCKT~PVLA NDDPLRPWIE RCLDLYGGLG RHPGLFPLFG
Novosphingobium sp. CCH12-A3        SVCKT~PVLA NDDPLRPWIE RCLDLYGGLG RHPGLFPLFG
Novosphingobium sp. B1              SVCKT~PVLA NDDPLRPWIE RCLDLYGGLG RHPGLFPLFG
Novosphingobium sp. AAP83           SVCKI~PVLA NDDPMRDWIE RSLDLFGGLG RHPGLFPLFG
Novosphingobium aromaticivorans     SVCKT~PVLA NDDPLRDWIE RCLDLYGGLG RHPGLFPLFG
Novosphingobium sp. B-7             SVCKT~PVLA NDDPLRDWLD RCLDLYGGLG RHPGLFPLFG
Novosphingobium sp. AAP1            SVCKT~PVLA NDDPLRDWLD RCLDLYGGLG RHPGLFPLFG
Altererythrobacter atlanticus(1)    SVCKTSPVLA DDDPLRDWIE RCLDLFGGLG RHPGLFPLFG
Erythrobacter sp. SG61-1L(1)        SVCQTSPVFA DDDPLRDWIE RCLDLYGGLG RHPGLFPLFG
Novosphingobium capsulatum          SVCKT~PVLA SDDPLRDWLD RCLDLYGGLG RHPGLFPLFG
Novosphingobium sp. SCN 66-18       SVCQT~PVLA EDDPLRDWIE RCLDLYGGLG RHPGLFPLFG
Novosphingobium sp. MD-1            SVCQT~PVLA EDDPLRDWIE RCLDLYGGLG RHPGLFPLFG
Altererythrobacter sp. 66-12        SVCKTSPVLA DDDPLRPWID NLLDMFGGLG RHPGLFPLFG
Sphingomonadales bacterium 63-6(1)  SVCQTSPVFA DDDPLRDWIE RCLDLYDGLG RHPGLFPLFG
Altererythrobacter sp. Root672(1)   SVCQNSPVLA EDDPLRPWIE RCLDLYDGLG RHPGLFPLFG
Sphingomonadales bacterium 63-6(2)  AVADT~PVLT EDDPLRDWIE RGFDLYGGLG RHPGLLPLFG
Erythrobacter sp. SG61-1L(2)        AVADT~PVLT DDDPLRDWID RGFDLYGGLG RHPGMLPLFG
Novosphingobium acidiphilum         SIADV~PPMT DDEPLRDWID RGFDLFGGLG RIPGMSPLFG
Novosphingobium sp. Fuku2-ISO-50    SVADI~PPMT DDEPLRDWID RGFDLYGGLG RIPGMSPLFG
Sphingobium sp. 66-54(1)            SVADT~PVLT DDDPLRDWIE RGFDLYGGLG RHPGLSPIFG
Altererythrobacter sp. Root672(2)   SVADI~PALP ADDPLRDWID RGFDLYGGLG RHPGMSPIFG
Altererythrobacter atlanticus(2)    SVCDT~PALA EDDPLRDWID RGFDLYGGIG RHPGLSNIFG
Novosphingobium sp. SCN 63-17       SVADT~PPMT EDDPLRDWID RGFDLYGGLG RIPGMSPLFG
Sphingobium sp. SCN 64-10           SVADT~PVLT DDDPLRDWIE RGFDLYGGLG RHPGLSPIFG
Sphingobium sp. SYK-6(2)            SVADT~PVLT DDDPLRDWIE RGFDLYGGLG RHPGLSPIFG
Novosphingobium sp. FSW06-99        SIADI~PPMT DDEPLRDWID RGFDLYGGLG RIPGMSPLFG
Sphingomonas hengshuiensis          SVATT~PPLT DDDPLRDWLD RGFDLYDGLG RHPGMHSLFG
Novosphingobium sp. PP1Y            SVATT~PPLT EDDPLRGWLD RGFDLYAGLG RHPGMHSLFG
Novosphingobium sp. ST904           SVATT~PPLT GDDPLRDWLD RGFDFFGGLG RHPAMHTLFG
Novosphingobium mathurense          SVATT~PPLT EDDPLRGWLD RGFDLHGGLG RHPGMHSLFG
Novosphingobium sp. KN65.2          SVATT~PPLT EDDPLRGWLD RGFDLHGGLG RHPGMHGLFG
Sphingobium sp. 66-54(2)            SVART~PPLT EDDPLRDWLD RGFDLFGGLG RHPAMHSLFG
Novosphingobium sp. MBES04          SVCTT~PPLT EDDPLRDWLD RGFDLYGGLG RHPGMHTLFG

                                              250        260        270        280
                                    ....|....| ....|....| ....|....| ....|....|
Sphingobium sp. SYK-6(1)            LKLREGDPEP FVRQTGPAGA GGQALNKG~P QTTKMPPRVA
Novosphingobium fuchskuhlense       LEQREGDPDL FNR~~~~~~~ ~~AGGQGG~I ~YKRNTGPAS
Novosphingobium sp. NDB2Meth1       LEEPEGGPAL FNR~~~~~~~ ~~AAGQGG~I ~YKRNTGPAS
Novosphingobium sp. AAP93           LEEPEGGPAL FNR~~~~~~~ ~~AAGQGG~I ~YKRNTGPAS
Novosphingobium subterraneum        LEQREGDPDL FNR~~~~~~~ ~~SAGQGG~I ~YKRNTGPDS
Novosphingobium sp. CCH12-A3        LEQREGDPDL FNR~~~~~~~ ~~SAGQGG~I ~YKRNTGPDS
Novosphingobium sp. B1              LEQREGDPDL FNR~~~~~~~ ~~AAGQGG~I ~YKRNTGPDS
Novosphingobium sp. AAP83           LEQREGDPDL FNR~~~~~~~ ~~AAGQGG~I ~YKRNTGPAS
Novosphingobium aromaticivorans     LEQREGDPDL FNR~~~~~~~ ~~AAGQGG~I ~YKRNTGPES
Novosphingobium sp. B-7             LVEREGDVPA FTR~~~~~~~ ~~QGGLGG~I ~YKRNTGPAS
Novosphingobium sp. AAP1            LVEREGDVPA FTR~~~~~~~ ~~QGGLGG~I ~YKRNTGPAS
Altererythrobacter atlanticus(1)    LPHPENGQEL FA~~~~~~~~ ~~PQGQGG~I ~HKRNTGVDS
Erythrobacter sp. SG61-1L(1)        LKQRDGDPEL FA~~~~~~~~ ~~PQGAGG~I ~HKRNTGPAS
Novosphingobium capsulatum          LVEREGDVPA FTR~~~~~~~ ~~QGGLGG~I ~YKRNTGPAS
Novosphingobium sp. SCN 66-18       LKQRDGDPDL FNR~~~~~~~ ~~QAGQGG~I ~YKRNTGPES
Novosphingobium sp. MD-1            LKQRDGDPDL FNR~~~~~~~ ~~QAGQGG~I ~YKRNTGPES
Altererythrobacter sp. 66-12        LKQREGDPPL FM~~~~~~~~ ~~PAGMGG~I ~HKRNTGVES
Sphingomonadales bacterium 63-6(1)  LKRREGDPEL FA~~~~~~~~ ~~PQGQGG~I ~HKRNTGPAS
Altererythrobacter sp. Root672(1)   LEQRPQDPPL FM~~~~~~~~ ~~PQGQGG~I ~HKRNTGPAS
Sphingomonadales bacterium 63-6(2)  LQLREGDPEP FQK~~~~~~~ ~~GPQMGG~L ~ARRNTGPAS
Erythrobacter sp. SG61-1L(2)        LQLREGDPEP FQK~~~~~~~ ~~GPQLGG~L ~ARRNTGPQS
```

```
Novosphingobium acidiphilum       LQPRPGDPEP FAK~~~~~~~ ~GPTLVGG~L ~ISRNTGPTS
Novosphingobium sp. Fuku2-ISO-50  LKLRAGDPEP FAK~~~~~~~ ~GPTLVGG~L ~VSRNTGPAS
Sphingobium sp. 66-54(1)          LQLREGDPEP FLK~~~~~~~ ~~GAGVGGGL ~YNRNTGPQS
Altererythrobacter sp. Root672(2) LPRRENDPEP FVR~~~~~~~ ~~DPQANG~L ~TKRNTGPQS
Altererythrobacter atlanticus(2)  LKQRAGDPDL FNR~~~~~~~ ~~NPMANG~L ~TSRNTGVKS
Novosphingobium sp. SCN 63-17     LKMREGDPEP FAK~~~~~~~ ~GPSLVAG~L ~VKRNTGPAS
Sphingobium sp. SCN 64-10         LQLREGDPEP FIK~~~~~~~ ~~GGAVGG~L ~ATRNTGPKS
Sphingobium sp. SYK-6(2)          LQLREGDPEP FIK~~~~~~~ ~~GGAVGG~L ~ATRNTGPKS
Novosphingobium sp. FSW06-99      LQVREGDPEP FAK~~~~~~~ ~GPTLVGG~L ~ISRNTGPAS
Sphingomonas hengshuiensis        LQLREGDPAP FLK~~~~~~~ AGIGTAPA~P ~VNRGAGTTP
Novosphingobium sp. PP1Y          LRLREGDPEP FAR~~~~~~~ DGAGIEVA~P ~VNRGTARSA
Novosphingobium sp. ST904         LRLRPGDPEP FES~~~~~~~ GGAGIEVA~P ~VNRGAAPQV
Novosphingobium mathurense        LQLREGDPEP FVR~~~~~~~ DGAGIEVA~P ~VNRGTARSA
Novosphingobium sp. KN65.2        LQLREGDPEP FVR~~~~~~~ DGAGIEVA~P ~VNRGTARSA
Sphingobium sp. 66-54(2)          LKLREGDPEP FVKQTGPAGA GGQGINKG~P QTTKMPPRGT
Novosphingobium sp. MBES04        LKLREGDPEP FDR~~~~~~~ TGLGIEPA~P ~VNQGSAEPA

                                            290         300
                                  ....|....| ....|....| .
Sphingobium sp. SYK-6(1)          EKAD------ ---------- -
Novosphingobium fuchskuhlense     TQAE~~~TQR ITEGMKK--- -
Novosphingobium sp. NDB2Meth1     TQAE~~~TQR ITEGMKK--- -
Novosphingobium sp. AAP93         TQAE~~~TQR ITEGMKK--- -
Novosphingobium subterraneum      TRAE~~~TQR ITEGMKK--- -
Novosphingobium sp. CCH12-A3      TRAE~~~TQR ITEGMKK--- -
Novosphingobium sp. B1            TRAE~~~TQR ITEGMKK--- -
Novosphingobium sp. AAP83         TSAE~~~TQR ITEGLKR--- -
Novosphingobium aromaticivorans   TRAE~~~TQR ITEGMKK--- -
Novosphingobium sp. B-7           TQAE~~~TQK ITQGMKQPA- -
Novosphingobium sp. AAP1          TQAE~~~TQK ITQGMKQPA- -
Altererythrobacter atlanticus(1)  TRAE~~~SEK ITQGMSKD-- -
Erythrobacter sp. SG61-1L(1)      TSTE~~~TQR ITEGMKAG-- -
Novosphingobium capsulatum        TQAE~~~TQK ITQGMKQPA- -
Novosphingobium sp. SCN 66-18     TRAE~~~TQR ITEGMAAA-- -
Novosphingobium sp. MD-1          TRAE~~~TQR ITEGMAAA-- -
Altererythrobacter sp. 66-12      TRAE~~~TRR ITEGMAKN-- -
Sphingomonadales bacterium 63-6(1) TSAE~~~TQR ITQGLAAD-- -
Altererythrobacter sp. Root672(1) TSAE~~~TQR ITEGMAKA-- -
Sphingomonadales bacterium 63-6(2) TSAE~~~TAM LKGSR----- -
Erythrobacter sp. SG61-1L(2)      TQAE~~~TAM MKGKS----- -
Novosphingobium acidiphilum       TAAE~~~TAK ITGREETAKN A
Novosphingobium sp. Fuku2-ISO-50  TSAE~~~TAK ITGRKEAANH A
Sphingobium sp. 66-54(1)          TAAE~~~TAR MKGEAAPAA- -
Altererythrobacter sp. Root672(2) TASE~~~TAR LKGEAASA-- -
Altererythrobacter atlanticus(2)  TAGE~~~TAR MKGQKAGA-- -
Novosphingobium sp. SCN 63-17     TAAE~~~TAH ITGKGEKAPA -
Sphingobium sp. SCN 64-10         TAAE~~~TAR LKGQTAPAN- -
Sphingobium sp. SYK-6(2)          TAAE~~~TAR LKGEKAPAA- -
Novosphingobium sp. FSW06-99      TSAE~~~TAK ITGRAEAASD A
Sphingomonas hengshuiensis        FPEPARAA-- ---------- -
Novosphingobium sp. PP1Y          EMAD------ ---------- -
Novosphingobium sp. ST904         EPAD------ ---------- -
Novosphingobium mathurense        EMAD------ ---------- -
Novosphingobium sp. KN65.2        EMAD------ ---------- -
Sphingobium sp. 66-54(2)          TKTD------ ---------- -
Novosphingobium sp. MBES04        TAS------- ---------- -
```

Supplementary Figure 7.16: Multiple sequence alignment of LigE homologs obtained from BLAST tool provided by NCBI where sequence of LigE from *Sphingobium* sp. SYK-6 was used as a template. Proteins analyzed are *Sphingobium* sp. SYK-6(1) (WP_014075192.1), *Novosphingobium fuchskuhlense* (WP_067913477.1), *Novosphingobium* sp. NDB2Meth1 (WP_072379148.1), *Novosphingobium* sp. AAP93 (WP_054121822.1), *Novosphingobium subterraneum* (WP_039335072.1), *Novosphingobium* sp. CCH12-A3 (WP_062343074.1), *Novosphingobium* sp. B1 (WP_084280065.1), *Novosphingobium* sp. AAP83 (WP_054106368.1), *Novosphingobium aromaticivorans* (WP_011446047.1), *Novosphingobium* sp. B-7 (WP_028657990.1), *Novosphingobium* sp. AAP1 (WP_054131978.1), *Altererythrobacter atlanticus*(1)

(WP_046903062.1), *Erythrobacter* sp. SG61-1L(1) (WP_054530882.1), *Novosphingobium capsulatum* (WP_062781915.1), *Novosphingobium* sp. SCN 66-18 (ODU67585.1), *Novosphingobium* sp. MD-1 (GAO55515.1), *Altererythrobacter* sp. 66-12 (OJU60283.1), *Sphingomonadales bacterium* 63-6(1) (OJW72302.1), *Altererythrobacter* sp. Root672(1) (WP_055920889.1), *Sphingomonadales bacterium* 63-6(2) (OJW69813.1), *Erythrobacter* sp. SG61-1L(2) (WP_054530343.1), *Novosphingobium acidiphilum* (WP_028641482.1), *Novosphingobium* sp. Fuku2-ISO-50 (WP_067743869.1), *Sphingobium* sp. 66-54(1) (OJY67521.1), *Altererythrobacter* sp. Root672(2) (WP_055921561.1), *Altererythrobacter atlanticus*(2) (WP_046903179.1), *Novosphingobium* sp. SCN 63-17 (ODU84760.1), *Sphingobium* sp. SCN 64-10 (ODT90752.1), *Sphingobium* sp. SYK-6(2) (WP_014077574.1), *Novosphingobium* sp. FSW06-99 (WP_067615430.1), *Sphingomonas hengshuiensis* (WP_044331491.1), *Novosphingobium* sp. PP1Y (WP_013832481.1), *Novosphingobium* sp. ST904 (WP_054436034.1), *Novosphingobium mathurense* (WP_079730390.1), *Novosphingobium* sp. KN65.2 (WP_054947729.1), *Sphingobium* sp. 66-54(2) (OJY68654.1), *Novosphingobium* sp. MBES04 (WP_039391125.1). Protein sequences were aligned using BioEdit tool.

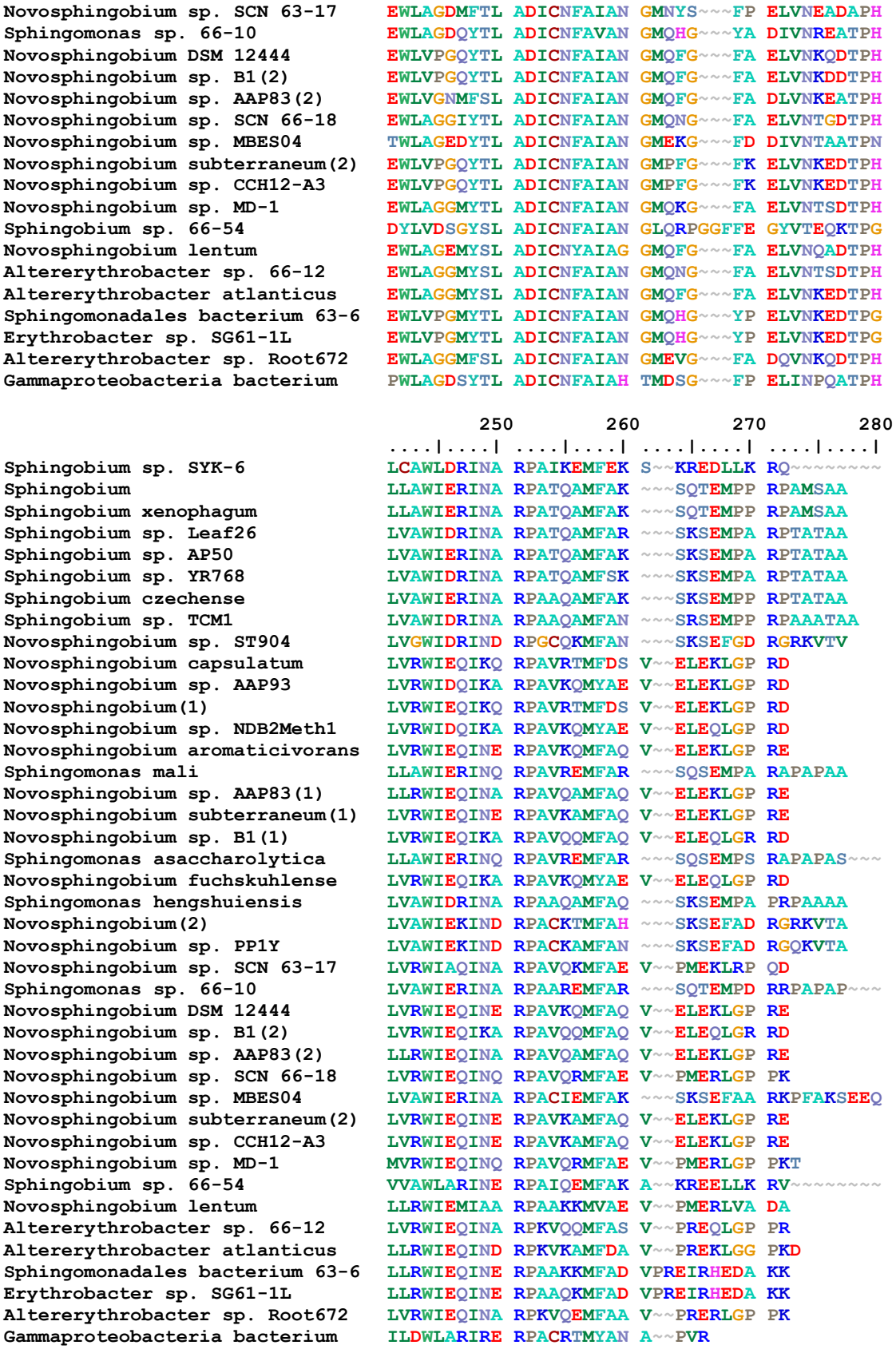
7.10 LigF homologs multiple sequence alignment

```
                                        10        20        30        40
                                ....|....| ....|....| ....|....| ....|....|
Sphingobium sp. SYK-6           M~~~~~~~~~ ~TLKLYSFGP GANSLKPLAT LYEK~~~GLE
Sphingobium                     ~~~~~~~~~~ ~MLTLYSFGP AANSMKPLLT LYEK~~~GLE
Sphingobium xenophagum          ~~~~~~~~~~ ~MLTLYSFGP AANSMKPLLT LYEK~~~GLE
Sphingobium sp. Leaf26          ~~~~~~~~~~ ~MLTLYSFGP AANSMKPLLA LYEK~~~GLE
Sphingobium sp. AP50            ~~~~~~~~~~ ~MLTLYSFGP AANSLKPLLA LYEK~~~GLE
Sphingobium sp. YR768           ~~~~~~~~~~ ~MLTLYSFGP AANSLKPLLA LYEK~~~GLE
Sphingobium czechense           ~~~~~~~~~~ ~MLTLYSFGP AANSLKPLLA LYEK~~~GLD
Sphingobium sp. TCM1            ~~~~~~~~~~ ~MLTLYSFGP AANSLKPLLA LYEK~~~GLD
Novosphingobium sp. ST904       ~~~~~~~~~~ ~MLTLYSFGP AANSLKPLLA LYEK~~~GLE
Novosphingobium capsulatum      ~~~~~~~~~~ ~MLKLYSFGP AANSMKPLLT LFEK~~~GLP
Novosphingobium sp. AAP93       ~~~~~~~~~~ ~MLKLYSFGP AANSMKPLLT LFEK~~~GLP
Novosphingobium(1)              ~~~~~~~~~~ ~MLKLYSFGP AANSMKPLLT LFEK~~~GLP
Novosphingobium sp. NDB2Meth1   ~~~~~~~~~~ ~MLKLYSFGP AANSMKPLLT LFEK~~~GLP
Novosphingobium aromaticivorans ~~~~~~~~~~ ~MLKLYSFGP AANSMKPLLT VFEK~~~GLD
Sphingomonas mali               ~~~~~~~~~~ ~MLTLYSFGP AANSLKPLLA LYEK~~~GLP
Novosphingobium sp. AAP83(1)    ~~~~~~~~~~ MTLKLYSFGP AANSMKPLLT VFEK~~~GLE
Novosphingobium subterraneum(1) ~~~~~~~~~~ MTLKLYSFGP AANSMKPLLT VFEK~~~GLD
Novosphingobium sp. B1(1)       ~~~~~~~~~~ MTLKLYSFGP AANSMKPLLT VFEK~~~GLD
Sphingomonas asaccharolytica    ~~~~~~~~~~ ~MLTLYSFGP AANSLKPLLA LYEK~~~GLP
Novosphingobium fuchskuhlense   ~~~~~~~~~~ ~MLKLYSFGP AANSLKPLLT LYEK~~~GLP
Sphingomonas hengshuiensis      ~~~~~~~~~~ ~MLTLYSFGP AANSLKPLLA LYEK~~~GLD
Novosphingobium(2)              ~~~~~~~~~~ ~MLTLYSFGP GANSLKPLLA LYEK~~~GLE
Novosphingobium sp. PP1Y        ~~~~~~~~~~ ~MLTLYSFGP GANSLKPLLA LYEK~~~GLE
Novosphingobium sp. SCN 63-17   ~~~~~~~~~~ ~MLKLYSFGP AANSMKPLLT LYEK~~~GTP
Sphingomonas sp. 66-10          ~~~~~~~~~~ ~MLTLYSFGP AANSLKPLLA LYEK~~~GLA
Novosphingobium DSM 12444       MVIPLGEDNT IMLKLYSFGP AANSMKPLLT VFEK~~~GLD
Novosphingobium sp. B1(2)       MVHPLGEDKT MTLKLYSFGP AANSMKPLLT VFEK~~~GLD
Novosphingobium sp. AAP83(2)    MVSPFGEDTI MTLKLYSFGP AANSMKPLLT VFEK~~~GLE
Novosphingobium sp. SCN 66-18   ~~~~~~~~~~ ~MLKLYSFGP AANSLKPLLT LYEK~~~GLP
Novosphingobium sp. MBES04      ~~~~~~~~~~ ~MLTLYSFGP GANSLKPLLA LYEK~~~GLE
Novosphingobium subterraneum(2) MVLPLGEDNA MTLKLYSFGP AANSMKPLLT VFEK~~~GLD
Novosphingobium sp. CCH12-A3    MVLPLGEDND MTLKLYSFGP AANSMKPLLT VFEK~~~GLD
Novosphingobium sp. MD-1        ~~~~~~~~~~ ~MLKLYSFGP GANSLKPLLT LYEK~~~GLP
Sphingobium sp. 66-54           M~~~~~~~~~ ~TLTLYSFGP GANSLKPLAT LYEK~~~GLE
Novosphingobium lentum          ~~~~~~~~~~ ~MLTLYSFGP GANSLKPLLT LYEK~~~GLE
Altererythrobacter sp. 66-12    ~~~~~~~~~~ ~MLTLYSFGP MANSLKPMLT LFEKFEFGRD
Altererythrobacter atlanticus   ~~~~~~~~~~ ~MLKLYSFGP GANSLKPMLT LYEK~~~GLE
Sphingomonadales bacterium 63-6 ~~~~~~~~~~ ~MLKLYSFGP GANSLKPMLT LHEK~~~GLD
Erythrobacter sp. SG61-1L       ~~~~~~~~~~ ~MLKLYSFGP GANSLKPMLT LHEK~~~GLD
Altererythrobacter sp. Root672  ~~~~~~~~~~ ~MLKLYSFGP GANSLKPMLT LFEK~~~GLE
Gammaproteobacteria bacterium   ~~~~~~~~~~ ~MLTLYSFGP SANSLKPLLT LYEK~~~GLD
```

```
                                          50          60          70          80
                                   ....|....| ....|....| ....|....| ....|....|
Sphingobium sp. SYK-6              FEQVFVDPSK FEQHSDWFKK INPRGQVPAL WH~~~~DGKV
Sphingobium                        FTPRFVDPTR FEHHEDWFKA INPRGQVPAL DH~~~~DGHI
Sphingobium xenophagum             FTPRFVDPTR FEHHEDWFKA INPRGQVPAL DH~~~~DGHI
Sphingobium sp. Leaf26             FVPRFVDPTR FEHHEDWFKA INPRGQVPAL DH~~~~DGHI
Sphingobium sp. AP50               FTPRFVDPTR FEHHEDWFKA INPRGQVPAL DH~~~~DGHI
Sphingobium sp. YR768              FTPRFVDPTR FEHHEDWFKA INPRGQVPAL DH~~~~DGHI
Sphingobium czechense              FTPRFVDPTR FEHHEDWFKA INPRGQVPAL DH~~~~DGHI
Sphingobium sp. TCM1               FTPRFVDPTR FEHHEDWFKA INPRGQVPAL DH~~~~DGHI
Novosphingobium sp. ST904          FTPRFVDPTR FEHHEDWFKK INPRGQVPAL DH~~~~DGHI
Novosphingobium capsulatum         FQANRLDPAK FEHHSDWFKA INPRGQVPAL VD~~~~GDKV
Novosphingobium sp. AAP93          FEKNRLDPAK FEHHTDWFKA INPRGQVPAL VD~~~~GEHV
Novosphingobium(1)                 FQANRLDPAK FEHHSDWFKA INPRGQVPAL VD~~~~GDKV
Novosphingobium sp. NDB2Meth1      FEKNRLDPAK FEHHTDWFKA INPRGQVPAL VD~~~~GEHV
Novosphingobium aromaticivorans    VEKHRLDPAK FEHHTDWFKA INPRGQVPAL VD~~~~GDKV
Sphingomonas mali                  FTPRFVDPRK FEHHEEWFKK INPRGQVPAL DH~~~~DGHI
Novosphingobium sp. AAP83(1)       VEKHRLDPAK FEHHTDWFKA INPRGQVPAL VD~~~~GGKV
Novosphingobium subterraneum(1)    VEKHRLDPAK FEHHTDWFKA INPRGQVPAL VDSGVDGDKV
Novosphingobium sp. B1(1)          VEKHRLDPAK FEHHTDWFKA INPRGQVPAL VD~~~~GEKV
Sphingomonas asaccharolytica       FTPRFIDPRK FEHHEEWFKK INPRGQVPAL DH~~~~DGHI
Novosphingobium fuchskuhlense      FEKNRLDPAK FEHHTDWFKA INPRGQVPAL VD~~~~GDHV
Sphingomonas hengshuiensis         FTPRFVDPTR FEHHEDWFKQ LNPRGQVPAL DH~~~~DGHI
Novosphingobium(2)                 FTPRFVDPTR FEHHEDWFKK INPRGQVPAL DH~~~~DGHI
Novosphingobium sp. PP1Y           FTPRFVDPTR FEHHEEWFKK INPRGQVPAL DH~~~~DGHI
Novosphingobium sp. SCN 63-17      FEGHRLNPAI FEHHQDWFKA INPRGQVPAL VD~~~~GDKV
Sphingomonas sp. 66-10             FTPRFIDPRR FEHHEDWFKR INPRGQVPAL DH~~~~DGHI
Novosphingobium DSM 12444          VEKHRLDPAK FEHHTDWFKA INPRGQVPAL VD~~~~GDKV
Novosphingobium sp. B1(2)          VEKHRLDPAK FEHHTDWFKA INPRGQVPAL VD~~~~GEKV
Novosphingobium sp. AAP83(2)       VEKHRLDPAK FEHHTDWFKA INPRGQVPAL VD~~~~GGKV
Novosphingobium sp. SCN 66-18      FEGHRLNPAI FEHHSDWFKA INPRGQVPAL VD~~~~RGKV
Novosphingobium sp. MBES04         FTPRFVDPTK FEHHEEWFKK INPRGQVPAL DH~~~~DGNV
Novosphingobium subterraneum(2)    VEKHRLDPAK FEHHTDWFKA INPRGQVPAL VDSGVDGDKV
Novosphingobium sp. CCH12-A3       VEKHRLDPAK FEHHTEWFKA INPRGQVPAL VDSGVDGDKV
Novosphingobium sp. MD-1           FEGHRLNPAT FEHHSDWFKA INPRGQVPAL VD~~~~EGRI
Sphingobium sp. 66-54              FTPRFVDPAK FEQHSDWYKK INPRGQVPAL DH~~~~DGKI
Novosphingobium lentum             FKGVRLNPAQ FEHHEDWFKA INPNGQVPAL VD~~~~NGRV
Altererythrobacter sp. 66-12       FTVHRLDPSK FEHHTDWFKA INPRGQVPAL KD~~~~GDRI
Altererythrobacter atlanticus      YEQHLLNPAK FEHHSDWFKE INPRGQVPAL WD~~~~DGKV
Sphingomonadales bacterium 63-6    YEQALLDPRK FEHHSDWFKK INPRGQVPAL ED~~~~RGHI
Erythrobacter sp. SG61-1L          YEQVLLDPRK FEHHSDWFKK INPRGQVPAL ED~~~~RGHI
Altererythrobacter sp. Root672     YEQHQLNPAK FEHHSDWYKA VNPRGQVPAL DD~~~~NGRI
Gammaproteobacteria bacterium      FESRYLNPAR FEQHEDWFKQ INPNAQAPAL DH~~~~DGRI

                                          90         100         110         120
                                   ....|....| ....|....| ....|....| ....|....|
Sphingobium sp. SYK-6              VTESTVICEY LEDVFPESGN ~SLRPADPFK RAEMRVWTKW
Sphingobium                        ITESTVICEY LEDAFPDA~P ~RLRPADPVG IAEMRVWTKW
Sphingobium xenophagum             ITESTVICEY LEDAFPDA~P ~RLRPVDPVG IAEMRVWTKW
Sphingobium sp. Leaf26             ITESTVICEY LEDAFPDA~P ~RLRPTDPVQ IAEMRVWTKW
Sphingobium sp. AP50               ITESTVICEY LEDAFPDA~P ~RLRPADPVQ IAEMRVWTKW
Sphingobium sp. YR768              ITESTVICEY LEDAFPDA~P ~RLRPVDPVQ IAEMRVWTKW
Sphingobium czechense              ITESTVICEY LEDAFPNA~T ~RLRPTDPVQ IAEMRVWTKW
Sphingobium sp. TCM1               ITESTVICEY LEDAFPDA~P ~RLRPTDPVQ IAEMRVWTKW
Novosphingobium sp. ST904          ITESTVICEY LEDAFPDA~P ~KLRPTDPVQ IAEMRVWTKW
Novosphingobium capsulatum         ITESTVICEY LEDEYPTQ~V ~SLRPADSYG RAQMRVWTKW
Novosphingobium sp. AAP93          ITESTVICEY LEDEYPTE~V ~KLRPDTSYG RAQMRVWTKW
Novosphingobium(1)                 ITESTVICEY LEDEYPTQ~V ~SLRPADSYG RAQMRVWTKW
Novosphingobium sp. NDB2Meth1      ITESTVICEY LEDEYPTE~V ~KLRPDTSYG RAQMRVWTKW
Novosphingobium aromaticivorans    VTESTVICEY LEDEYPTE~V ~ALRPADSFG KAQMRIWTKW
Sphingomonas mali                  ITESTVICEY LEDAFPDA~P ~RLRPTDPIK IAEMRVWTKW
Novosphingobium sp. AAP83(1)       ITESTVICEY LEDEYPTE~V ~KLRPADSYG RAQMRVWTKW
Novosphingobium subterraneum(1)    ITESTVICEY LEDEYPTE~V ~SLRPADSFG RAQMRIWTKW
Novosphingobium sp. B1(1)          VTESTVICEY LEDEYPTE~V ~SLRPADSFG RAQMRIWTKW
Sphingomonas asaccharolytica       ITESTVICEY LEDAFPDA~P ~RLRPADPVA TAEMRVWTKW
Novosphingobium fuchskuhlense      ITESTVICEY LEDEYPTE~V ~KLRPETSWG KAQMRVWTKW
Sphingomonas hengshuiensis         ITESTVICEY LEDAFPDC~A IKLRPADPIR IAEMRVWTKW
```

```
Novosphingobium(2)                      ITESTVICEY LEDAFPEA~P ~RLRPVDPVN IAEMRVWTKW
Novosphingobium sp. PP1Y                ITESTVICEY LEDAFPEA~P ~RLRPVDPVM IAEMRVWTKW
Novosphingobium sp. SCN 63-17           ITESTVICEY LEDEHPGE~V ~KLRPADSYG RAQMRIWTKW
Sphingomonas sp. 66-10                  VTESTVICEY LEDAFPDA~P ~RLRPTDPLA IAEMRVWTKW
Novosphingobium DSM 12444               VTESTVICEY LEDEYPTE~V ~ALRPADSFG KAQMRIWTKW
Novosphingobium sp. B1(2)               VTESTVICEY LEDEYPTE~V ~SLRPADSFG RAQMRIWTKW
Novosphingobium sp. AAP83(2)            ITESTVICEY LEDEYPTE~V ~KLRPADSYG RAQMRVWTKW
Novosphingobium sp. SCN 66-18           VTESTVICEY LEDEHPTE~V ~ALRPADSFG KAEMRVWTKW
Novosphingobium sp. MBES04              ITESTVICEY LEDAFPDA~P ~RLRPTDPVQ IAEMRVWTKW
Novosphingobium subterraneum(2)         ITESTVICEY LEDEYPTE~V ~SLRPADSFG RAQMRIWTKW
Novosphingobium sp. CCH12-A3            ITESTVICEY LEDEYPTE~V ~SLRPADSFG RAQMRIWTKW
Novosphingobium sp. MD-1                VTESTVICEY LEDAYPTE~V ~ALRPADSFG KAEMRVWTKW
Sphingobium sp. 66-54                   VTESTVICEY LEDVFPDAGT ~SLRPADPYK RAEMRVWTKW
Novosphingobium lentum                  ITESTVICEY LEDEYPTA~V ~KLRPADSYG RAQMRVWTKW
Altererythrobacter sp. 66-12            VTESTVICEY LEDAHPTA~V ~KLRPDDPYD RAQMRVWTKW
Altererythrobacter atlanticus           VTESTVICEY LEDAHPTE~V ~KLRPEDPFD RAQMRIWTKW
Sphingomonadales bacterium 63-6         ITESTVICEY LEDEWPTD~V ~KLRPDTSWE RADMRVWTKW
Erythrobacter sp. SG61-1L               ITESTVICEY LEDEWPTD~V ~KLRPDTSWE RADMRVWTKW
Altererythrobacter sp. Root672          VTESTVICEY LEDAHPTA~V ~KLRPDDPYD RAQMRVWTKW
Gammaproteobacteria bacterium           ITESTVICEY LEDLFPGT~P ~RLRPEDPYQ RAQMRIWTKW

                                                130        140        150        160
                                        ....|....| ....|....| ....|....| ....|....|
Sphingobium sp. SYK-6                   VDEYFCWCVS TIGWAFGIKA IAQKMSDEEF EEHINKNVPI
Sphingobium                             VDEYFCWCVS TIGWERMIGP MARKLSDEEF EEKL~KHIPI
Sphingobium xenophagum                  VDEYFCWCVS TIGWERMIGP MARKLSDEEF EEKL~KHIPI
Sphingobium sp. Leaf26                  VDEYFCWCVS TIGWERMIGP MARKLSDAEF EEKL~KHIPI
Sphingobium sp. AP50                    VDEYFCWCVS TIGWERMIGP MARKLSDAEF EEKL~KHIPV
Sphingobium sp. YR768                   VDEYFCWCVS TIGWERMIGP MARKLSDAEF EEKL~KHIPI
Sphingobium czechense                   VDEYFCWCVS TIGWERMIGP MARKLSDAEF EEKL~KHIPI
Sphingobium sp. TCM1                    VDEYFCWCVS TIGWERMIGP MARKLSDEEF EEKL~KHIPI
Novosphingobium sp. ST904               VDEYFCWCVS TLGWERGIGP MARALSDEEF EAKV~ARIPI
Novosphingobium capsulatum              VDEYFCWCVS TIGWHRYVGK MVQGLSDQEF EDKV~KAIPV
Novosphingobium sp. AAP93               VDEYFCWCVS TIGWHRYVGN MVKGLSDAEF EEKV~KAIPV
Novosphingobium(1)                      VDEYFCWCVS TIGWHRYVGK MVQGLSDQEF EDKV~KAIPV
Novosphingobium sp. NDB2Meth1           VDEYFCWCVS TIGWHRYVGN MVKGLSDAEF EEKV~KAIPV
Novosphingobium aromaticivorans         VDEYFCWCVS TIGWHRYVGN MVKSLSDAEF EEKV~KAIPV
Sphingomonas mali                       VDEYFCWCVS TIGWERMIGP MARALSDAEF EEQV~KRIPV
Novosphingobium sp. AAP83(1)            VDEYFCWCVS TIGWHRYVGN MVKALSDAEF EEKV~AAIPV
Novosphingobium subterraneum(1)         VDEYFCWCVS TIGWHRYVGN MVKGLTDAEF EEKV~KAIPV
Novosphingobium sp. B1(1)               VDEYFCWCVS TIGWHRYVGN MVKGLTDAEF EAKV~AAIPV
Sphingomonas asaccharolytica            VDEYFCWCVS TIGWERMIGP MARALSDSEF EEQL~ARIPV
Novosphingobium fuchskuhlense           VDEYFCWCVS TIGWHRYVGN MVKSLSDAEF EEKV~KAIPV
Sphingomonas hengshuiensis              VDEYFCWCVS TIGWERMIGP MARAYSDAEF EEKM~ARIPI
Novosphingobium(2)                      VDEYFCWCVS TIGWERMIGP MARALSDEEF EAKV~ARIPV
Novosphingobium sp. PP1Y                VDEYFCWCVS TIGWERMIGP MARALSDEEF EAKV~ARIPV
Novosphingobium sp. SCN 63-17           VDEYFCWCVS TIGWHRYVGN MVKSLSDEAF EEKV~KAIPV
Sphingomonas sp. 66-10                  VDEYFCWCVS TIGWERMIGP MARALSDEEF EEKV~KRIPV
Novosphingobium DSM 12444               VDEYFCWCVS TIGWHRYVGN MVKSLSDAEF EEKV~KAIPV
Novosphingobium sp. B1(2)               VDEYFCWCVS TIGWHRYVGN MVKGLTDAEF EAKV~AAIPV
Novosphingobium sp. AAP83(2)            VDEYFCWCVS TIGWHRYVGN MVKALSDAEF EEKV~AAIPV
Novosphingobium sp. SCN 66-18           VDEYFCWCVS TIGWHRGVSH MAQRLSDAEF EEHL~KKIPI
Novosphingobium sp. MBES04              VDEYFCWCVS TIGWERGIGP MARALSDEEF EEKV~KRIPI
Novosphingobium subterraneum(2)         VDEYFCWCVS TIGWHRYVGN MVKGLTDAEF EEKV~KAIPV
Novosphingobium sp. CCH12-A3            VDEYFCWCVS TIGWHRYVGN MVKGLTDAEF EEKV~KAIPV
Novosphingobium sp. MD-1                VDEYFCWCVS TIGWHRGVSH MAQRLSDAEF EEHL~KKIPV
Sphingobium sp. 66-54                   VDEYFCWCVS TIGWAFGIKA IAQKMSDEEF EEHIRKNVPI
Novosphingobium lentum                  VDEYFCWCVS TIGWSRMISG MARAISAEEF EEKI~KRIPI
Altererythrobacter sp. 66-12            VDEYFCWCVS TIGWHRGVSH MAKALSDQEF EEHL~KKIPI
Altererythrobacter atlanticus           VDEYFCWCVS TIGWHRGVRQ MAQQLSDDEF EEHL~KKIPI
Sphingomonadales bacterium 63-6         VDEYFCWCVS TIGWHRYVGN MVKGLSDEEF EEKV~KNIPV
Erythrobacter sp. SG61-1L               VDEYFCWCVS TIGWHRYVGN MVKGLSDEEF EEKV~KNIPV
Altererythrobacter sp. Root672          VDEYFCWCVS TIGWHRGVRF MAQQLSDAEF EEHL~KKIPI
Gammaproteobacteria bacterium           VDEYFCWCVS TIGWERRISG IARSVSEEEF ERLV~ARIPL

                                                170        180        190        200
                                        ....|....| ....|....| ....|....| ....|....|
```

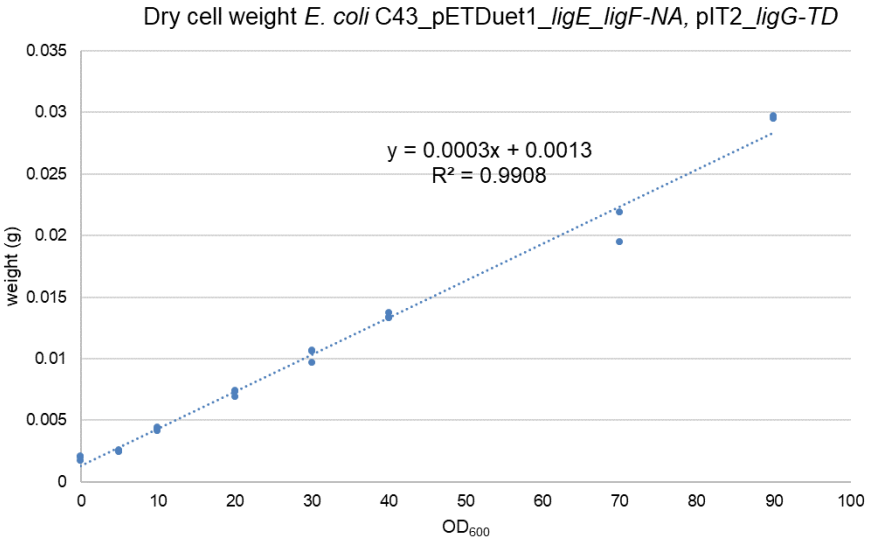
```
Sphingobium sp. SYK-6                PEQQLKWRRA RNGFPQEMLD EEFRKVGVSV ARLEETLSKQ
Sphingobium                          PEQQAKWRNA RAGFPQAVLD EEMRKIRVSI DRLEKRLAQS
Sphingobium xenophagum               PEQQAKWRNA RAGFPQAVLD EEMRKIRVSI DRLEKRLAQS
Sphingobium sp. Leaf26               PEQQAKWRSA RAGFPQAVLD EEMRKIRVSI DRLEQRLSQS
Sphingobium sp. AP50                 PEQQAKWRSA RAGFPQAVLD EEMRKIRVSI DRLEQRLSQS
Sphingobium sp. YR768                PEQQAKWRSA RAGFPQAVLD EEMRKIRVSI DRLEKRLSQS
Sphingobium czechense                PEQRAKWRSA RAGFPQAVLD EEMRKIRVSI DRLEQRLAQS
Sphingobium sp. TCM1                 PEQQAKWRSA RAGFPQAVLD EEMRKIRVSI ERLEKRLSQS
Novosphingobium sp. ST904            PEQRTKWRTA RAGFPRDVLE EEMRKIRVSV DRVEKRLSES
Novosphingobium capsulatum           VEQQVKWRRA REGFPQDLLD EEMRKIAFSV RRLDDHLRDH
Novosphingobium sp. AAP93            VEQQVKWRRA REGFPQDMLD EEMRKIGYSV ERLDAHLRQH
Novosphingobium(1)                   VEQQVKWRRA REGFPQDLLD EEMRKIAFSV RRLDDHLRDH
Novosphingobium sp. NDB2Meth1        VEQQVKWRRA REGFPQDMLD EEMRKIAYSV ERLDAHLREH
Novosphingobium aromaticivorans      IEQQVKWRRA REGFPQDMLD EEMRKIAYSV RKLDDHLADH
Sphingomonas mali                    PEQQVKWRNA RKGFPKELLD EEMRKVRVSV DKLEKRLAES
Novosphingobium sp. AAP83(1)         VEQQVKWRRA REGFPQDMLE EEMRKIAFSV RRLDDHLADN
Novosphingobium subterraneum(1)      VEQQVKWRRA REGFPQDMLD EEMRKIGYSV RKLDDHLADH
Novosphingobium sp. B1(1)            VEQQVKWRRA REGFPQDMLD EEMRKIAYSV RKLDDHLADH
Sphingomonas asaccharolytica         AEQKVKWRNA RNGFPKEVLD EEMRKIGVSV DRLEKRLSQS
Novosphingobium fuchskuhlense        VEQQVKWRRA REGFPQDLLD EEMRKIGYSV ERLDAHLRQH
Sphingomonas hengshuiensis           PEQQAKWRAA RAGFPKATLD EEMRKIGVSV AKLEARLAES
Novosphingobium(2)                   PEQRTKWRTA RAGFPKDVLA EEMRKIGVSV NRLEMRLAES
Novosphingobium sp. PP1Y             PEQRTKWRTA RTGFPKEVLD EEMRKIGVSV NRLETRLAES
Novosphingobium sp. SCN 63-17        FEQQVKWRRA REGFPQDLLD EEMRKIAFSV TRLNDHLADH
Sphingomonas sp. 66-10               VEQQVKWRNA RAGFPKEVLD EEMRKIRFSI DRLEKRLSES
Novosphingobium DSM 12444            IEQQVKWRRA REGFPQDMLD EEMRKIAYSV RKLDDHLADH
Novosphingobium sp. B1(2)            VEQQVKWRRA REGFPQDMLD EEMRKIAYSV RKLDDHLADH
Novosphingobium sp. AAP83(2)         VEQQVKWRRA REGFPQDMLE EEMRKIAFSV RRLDDHLADN
Novosphingobium sp. SCN 66-18        PEQQVKWRRA REGFPQDLLD EEMRKIAVSV RRLDDHLADH
Novosphingobium sp. MBES04           PEQQAKWRSA RAGFPKEVLD EEMRKIRVSI DRLEKRLSES
Novosphingobium subterraneum(2)      VEQQVKWRRA REGFPQDMLD EEMRKIGYSV RKLDDHLADH
Novosphingobium sp. CCH12-A3         VEQQVKWRRA REGFPQDMLD EEMRKIGYSV RKLDDHLANH
Novosphingobium sp. MD-1             PEQQVKWRRA REGFPQDLLD EEMRKIAVSV RRLDDHLADN
Sphingobium sp. 66-54                PEQQLKWRRA RNGFPQEMLD EEYRKVGVSI ERLEATLSRQ
Novosphingobium lentum               PEQQVKWRRA RDGFPQDMLD EEMRKIGVSV RRLNDHLADH
Altererythrobacter sp. 66-12         PEQQVKWRRA REGFPQDLLD EEMRKIAVSV RKLDDHLADN
Altererythrobacter atlanticus        PEQQVKWRRA REGFPQDLLD EEMRKIGVSV RKLDDHLADN
Sphingomonadales bacterium 63-6      VEQQVKWRRA REGFPQEMLD EEMRKIAYSV RKLDDHLRDH
Erythrobacter sp. SG61-1L            VEQQVKWRRA REGFPQEMLD EEMRKIAYSV RKLDDHLRDN
Altererythrobacter sp. Root672       PEQQVKWRRA REGFPQDLLD EEMRKIGVSV RKLDDHLADN
Gammaproteobacteria bacterium        KEQQLKWRNA RNGFDKKVLE EEMRKIDYSV KKLEARLSQS

                                              210        220        230        240
                                     ....|....| ....|....| ....|....| ....|....|
Sphingobium sp. SYK-6                DYLVDTGYSL ADICNFAIAN GLQRPGGFFG DYVNQEKTPG
Sphingobium                          PWLAGDEYTL ADICNFAIAN GMQFG~~~HA DIVNGQATPH
Sphingobium xenophagum               PWLAGDEYTL ADICNFAIAN GMQFG~~~HA DIVNGQATPH
Sphingobium sp. Leaf26               PWLAGEDFTL ADICNFAIAN GMQFG~~~YA DIVNKDATPH
Sphingobium sp. AP50                 LWLAGEDYTL ADICNFAIAN GMQFG~~~YA DIVNKDMTPH
Sphingobium sp. YR768                PWLAGEDYTL ADICNFAIAN GMQFG~~~YS DIVNKTATPH
Sphingobium czechense                PWLAGDDYTL ADVCNFAIAN GMQFG~~~YA DIVNREATPH
Sphingobium sp. TCM1                 PWLAGEVYTL ADICNFAIAN GMQFG~~~YA DIVNPDATPN
Novosphingobium sp. ST904            TWLVGEQFTL ADVCNFAIAN GMQNS~~~FA DIVNREATPH
Novosphingobium capsulatum           EWLVPGQYTL ADICNFAIAN GMQFG~~~FA ELVNREDTPH
Novosphingobium sp. AAP93            EWLVPGQYTL ADICNFAIAN GMQVG~~~FA ELVNKEKTPG
Novosphingobium(1)                   EWLVPEQYTL ADICNFAIAN GMQFG~~~FA ELVNREDTPH
Novosphingobium sp. NDB2Meth1        EWLVPGQYTL ADICNFAIAN GMQVG~~~FA ELVNKEKTPG
Novosphingobium aromaticivorans      EWLVPGQYTL ADICNFAIAN GMQFG~~~FA ELVNKQDTPH
Sphingomonas mali                    EWLAGDDYTL ADICNFAIAN GMQHG~~~FA EIVNREASPH
Novosphingobium sp. AAP83(1)         EWLVGNMFSL ADICNFAIAN GMQFG~~~FA DLVNKEATPH
Novosphingobium subterraneum(1)      EWLVPGQYTL ADICNFAIAN GMPFG~~~FK ELVNKEDTPH
Novosphingobium sp. B1(1)            EWLVPGQYTL ADICNFAIAN GMQFG~~~FA ELVNKDDTPH
Sphingomonas asaccharolytica         PWLAGDDYTL ADICNFAIAN GMQHG~~~FA EIVNREASPH
Novosphingobium fuchskuhlense        EWLVPGQYTL ADICNFAIAN GMQLG~~~FP EFVSAEKTPG
Sphingomonas hengshuiensis           EWLAGPDYTL ADICNFAIAN RMQHG~~~FA ELVNTAATPH
Novosphingobium(2)                   PWLAGDEFTL ADVCNFAIAN GMQNG~~~FS DIVNREATPH
Novosphingobium sp. PP1Y             PWLAGENFSL ADVCNFAIAN GMQNG~~~FS DIVNREATPH
```



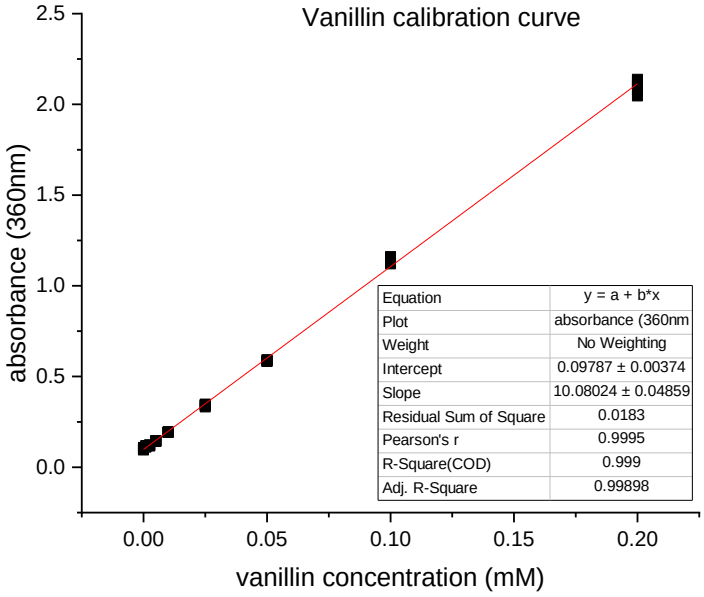
Supplementary Figure 7.17: Multiple sequence alignment of LigF homologs obtained from BLAST tool provided by NCBI where sequence of LigF from *Sphingobium* sp. SYK-6 was used as a template. Proteins analyzed are *Sphingobium* sp. SYK-6 (WP_014075191.1), *Sphingobium* (WP_017182419.1, multispecies

family protein), *Sphingobium xenophagum* (WP_019052363.1), *Sphingobium* sp. Leaf26 (WP_056685155.1), *Sphingobium* sp. AP50 (SEI61226.1), *Sphingobium* sp. YR768 (SER63234.1), *Sphingobium czechense* (WP_066609183.1), *Sphingobium* sp. TCM1 (WP_066854755.1), *Novosphingobium* sp. ST904 (WP_054436033.1), *Novosphingobium capsulatum* (WP_062782323.1), *Novosphingobium* sp. AAP93 (WP_054121580.1), *Novosphingobium*(1) (WP_022675760.1, multispecies family protein), *Novosphingobium* sp. NDB2Meth1 (WP_072381855.1), *Novosphingobium aromaticivorans* (WP_041551020.1), *Sphingomonas mali* (WP_066818898.1), *Novosphingobium* sp. AAP83(1) (KPF91204.1), *Novosphingobium subterraneum*(1) (KHS49048.1), *Novosphingobium* sp. B1(1) (SMC30538.1), *Sphingomonas asaccharolytica* (WP_066804606.1), *Novosphingobium fuchskuhlense* (WP_067910024.1), *Sphingomonas hengshuiensis* (WP_044331490.1), *Novosphingobium*(2) (WP_054947728.1, multispecies family protein), *Novosphingobium* sp. PP1Y (WP_013832480.1), *Novosphingobium* sp. SCN 63-17 (ODU83729.1), *Sphingomonas* sp. 66-10 (OJU17901.1), *Novosphingobium aromaticivorans* DSM 12444 (ABD26530.1), *Novosphingobium* sp. B1(2) (WP_084276739.1), *Novosphingobium* sp. AAP83(2) (WP_082356151.1), *Novosphingobium* sp. SCN 66-18 (ODU70795.1), *Novosphingobium* sp. MBES04 (WP_039391123.1), *Novosphingobium subterraneum*(2) (WP_082013229.1), *Novosphingobium* sp. CCH12-A3 (WP_082734789.1), *Novosphingobium* sp. MD-1 (GAO55900.1), *Sphingobium* sp. 66-54 (OJY68655.1), *Novosphingobium lentum* (WP_068075965.1), *Altererythrobacter* sp. 66-12 (OJU59935.1), *Altererythrobacter atlanticus* (WP_046904735.1), *Sphingomonadales bacterium* 63-6 (OJW61742.1), *Erythrobacter* sp. SG61-1L (WP_054529921.1), *Altererythrobacter* sp. Root672 (WP_055919008.1), *Gammaproteobacteria bacterium* RIFCSPLOWO2_02_FULL_56_15 (OGT78215.1). Protein sequences were aligned using BioEdit tool.

7.11 Calibration curves



Supplementary Figure 7.18: Calibration curve of the dry cell weight of *E. coli* C43.



Supplementary Figure 7.19: Calibration curve of the absorbance of the vanillin.

ACKNOWLEDGEMENTS

First and foremost, I would like to give a special thanks to Prof. Dr. Anett Schallmey for not only allowing me to work in her laboratories under her supervision, but for also being an endless source of knowledge and support. Your scientific mind and constant creativity are a never-ending motivation for me. You not only supervised my PhD work, but you taught me how to overcome unexpected situations and how to keep calm in stressful circumstances, which made me a better scientist and better person, and for that I am forever grateful.

I would like to thank Prof. Dr. Dieter Jahn for being part of my PhD defence as a second reviewer and for his valuable time spent reading my work.

I would also like to thank Prof. Dr. Ralf Mendel for being the chairman of the examination committee.

This research was made possible due to the funding from the “Niedersächsisches Ministerium für Wissenschaft und Kultur”, which I am thankful for.

The long and exhausting PhD journey always felt satisfying and joyful because of the wonderful and smart colleagues that I met during my research. Melanie, your ability to listen and give heartful advice turned my most miserable days into wonderful ones. Honestly, you are the kindest person I know, and I am extremely thankful to have met you. I want to thank Julia for her words of support when I was down and the introduction to the “world of PhD”; Willem for his wit, biochemical expertise and our shared sense of humor; Patrick for his help with MS analysis and enduring optimistic nature; Marcus for his organizational efforts and valuable molecular biology input; Elia for his great personality and shared ability to laugh at our own laboratory mistakes; Jenny for her very kind nature and always being reliable; Rita for inspirational dedication and willingness to continue with several unfinished parts of this project; Hazel and Kathi for endless discussions about science, cats and politics while drinking coffee (at 10am or 3pm) and Jhon Alex for his interaction knowledge and strong socializing skills. Separately, I want to say thank you to Hauke, who shared the lignin-themed project with me. I am very grateful for our discussions and willingness to help each other, trying to combine our knowledge and expertise to overcome the challenges connected with our project. Thank you very much! From my colleagues, I want to individually thank Laura, the best flatmate, the ideal person to complain to and the great confidence booster. Your easy-going personality, knowledge, and ability to see problems from a wide perspective allowed me to not take myself too seriously, but still always try my best. Our late WhatsApp chats kept me sane and I always ended up in a better mood. You truly are not just an

exceptional colleague but a wonderful friend. Thank you!

Osobitne sa chcem poďakovať mojim rodičom, Janke a Mariánovi. Ďakujem Vám za všetko, čo ste počas môjho študentského života pre mňa obetovali, že ste ma podporovali aj keď to určite bolo náročné. Vaša výchova, obetavosť a ochota zo mňa spravili človeka, ktorým som dnes. Ďakujem! Mojim súrodencom Ninke, Emke a Gabkovi sa chcem poďakovať za neprestajnú oporu a ochotu počúvať moje neustále sťažnosti, ste super! Aj keď nikto z mojej rodiny netuší, o čom táto práca vlastne je, bez ich podpory by táto publikácia nevznikla, a ja som mimoriadne vďačná za to, že ich mám.

At last but not least, I want to say a huge thanks to Matt. You were with me on this “run” from the very beginning, and sometimes I am even surprised that you have stayed with me until now. Without your constant support I would probably not have started my PhD and I would definitely not have finished it. Your patience, calm nature and open mind created the necessary “buffer” for all of my tempers, anxieties and thoughts of giving up. My PhD run is at the end and I am happy that I have ran it with you by my side. I am looking forward to the many shared marathons that lie ahead.

Academic CV

Personal information

Name:	Jana Husarčíková
Birthday:	03. March 1991
Place of birth:	Ružomberok, Slovakia
Nationality:	Slovak

Employment history

PhD Biochemistry – Technische Universität Braunschweig, Institute for Biochemistry, Biotechnology and Bioinformatics (Aug 2015 – Sept 2019)

Thesis title: Mechanistic and biocatalytic investigation of glutathione-dependent, ligninolytic enzymes

Education history

MSc Biotechnology (Ing)

Slovak University of Technology in Bratislava (2013 – 2015)

Master thesis: “Preparation and Applications of Recombinant Decarboxylases in Biocatalysis”

Grade: A (excellent)

BSc Biotechnology and Food technology (Bc)

Slovak University of Technology in Bratislava (2010 – 2013)

Bachelor thesis: “Biotransformácie pomocou monoamín oxidázových (MAO) enzýmov” (Biotransformations with Monoamine Oxidase (MAO))

Grade: A (excellent)

Vorveröffentlichungen

Publikationen

Jana Husarcikova, Hauke Voß, Pablo Domínguez de María and Anett Schallmey. Microbial β-etherases and glutathione lyases for lignin valorization in biorefineries: current state and future perspectives, Appl. Microbiol. Biotechnol. **2018**, *102*, 5391−5401. https://doi.org/10.1007/s00253-018-9040-3.

Kristina Markosova, Jana Husarcikova, Monika Halasova, Robert Kourist, Martin Rosenberg, Radek Stloukal, Ludmila Zajoncova, Martin Rebros. Immobilization of Arylmalonate Decarboxylase, Catalysts. **2018**, 8, *603*.
https://doi.org/10.3390/catal8120603

Jana Husarcikova and Anett Schallmey. Whole-cell cascade for the preparation of enantiopure β-*O*-4 aryl ether compounds with glutathione recycling, J. Biotechnol. **2019**, *293*, 1-7. https://doi.org/10.1016/j.jbiotec.2019.01.002.

Tagungsbeiträge

Jana Husarcikova and Anett Schallmey. Lignin valorization with GSH-dependent enzymes. 2018, 9th International Congress on Biocatalysis, Hamburg, Germany.

Posterbeiträge

Jana Husarcikova, Pere Picart, Hauke Voß and Anett Schallmey. Biomass valorization using GSH-dependent enzymes. 2016, The EMBO conference: "The biochemistry and chemistry of biocatalysis: From understanding to design", Oulu, Finland.

Jana Husarcikova, Hauke Voß and Anett Schallmey. Biomass valorization using glutathione-dependent enzymes. Biotrans 2017, 13th International Symposium on Biocatalysis and Biotransformations, Budapest, Hungary.

Jana Husarcikova and Anett Schallmey. Novel insights into β-etherase catalysis. 2018, 8th International Cebitec Research Conference, Bielefeld, Germany

Jana Husarcikova and Anett Schallmey. Novel insights into β-etherase catalysis. Biocat 2018, 9th International Congress on Biocatalysis, Hamburg, Germany.

Jana Husarcikova and Anett Schallmey. Whole-cell cascade using β-etherases and glutathione lyase with intracellular glutathione supply and recycling. Biotrans 2019, 14th International Symposium on Biocatalysis and Biotransformations, Groningen, Netherlands.